Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 39, NUMBER 5 MAY, 1949

A New Classification System for Radiation Detectors*

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts
(Received February 4, 1949)

In Part I, a classification system for radiation detectors is proposed. The system is based upon the manner in which the noise equivalent power depends upon the time constant and the sensitive area, when the sensitivity is limited either by radiation fluctuations or by internally generated noise.

The major part of Part I is devoted to establishing the reference condition for the measuring of the noise equivalent power. In establishing the reference condition of measurement (Section 6), the output of the detector in the absence of a signal is first equalized so that the noise power per unit frequency band width is constant. The reference time constant is then defined in terms of the relative response as a function of signal modulation frequency (Section 7). The equalization is then modified by the addition of an *RC* low pass filter with a time constant equal to the reference time constant, and the noise equivalent power is measured. The power so obtained is termed the noise equivalent power in reference condition *A*, denoted by P_m. This procedure has the property that the noise equivalent power of the detector is measured in the presence of noise whose band width is equal to the band width of the detector.

In order to establish the reference condition of measurement it is necessary to consider the various sources of noise which are involved in detectors (Section 2), and the various types of time constants (Section 3). The important concept of responsivity-to-noise ratio is defined in Section 4. A general theorem involving the sensitive area and the responsivity-to-noise ratio is established in Section 5.

After the question of detectors with non-uniform spectral sensitivity is discussed in Section 8, the classification system is defined in Section 9: A detector is a Type *I* detector if its noise equivalent power depends upon its sensitive area *A* and its reference time constant τ in accordance with

$$P_m = A^{\frac{1}{2}}/k_1\tau^{\frac{1}{2}},$$

where k_1 is independent of *A* and τ. A Type *II* detector is defined by

$$P_m = A^{\frac{1}{2}}/k_2\tau,$$

and more generally, a Type *n* detector is defined by

$$P_m = A^{\frac{1}{2}}/k_n\tau^{\frac{1}{n}}.$$

The usefulness of the proposed classification is illustrated in Section 10 by its application to sequential scanning systems.

In Part II the classification system proposed in Part I is used to determine the type number of each of eight different kinds of detectors. By a detailed analysis of each kind of detector, it is found that all of the detectors studied are either Type *I* or Type *II* detectors. The results of the analysis are summarized in Table I.

The detectors studied include bolometers (Section 2), thermocouples and thermopiles (Section 3), the Golay detector (Section 4), photographic plates and films (Section 5), vacuum and gas photo-tubes and photo-multiplier tubes (Section 6), and dipole antennas (Section 7). In the case of the dipole antenna it is shown that Johnson noise becomes equivalent to the noise produced by radiation fluctuations.

PART I. THE CLASSIFICATION SYSTEM

1. Introduction

AN ideal radiation detector might be defined as one which counts every one of the incident quanta which constitute the signal to be detected, and which gives no response when this signal is absent. In general, no detector has this property, although some detectors for high energy photons, such as the proportional counter, are nearly ideal in this respect.

The difficulty is that all detectors have a background response of some kind. The zero-frequency component of the background response may be eliminated by a subtraction method, but the fluctuations in this background remain. No signal can be detected unless the amplitude of its response is comparable with or greater than the amplitude of the fluctuations which lie in the frequency band of the signal.

* This is the second of a series of papers on radiation detectors, the second and third of which appear in this issue: I. "The ultimate sensitivity of radiation detectors," J. Opt. Soc. Am. **37**, 879–890 (1947), referred to as Paper I.
II. This paper, referred to as Paper II.
III. "Factors of merit for radiation detectors," in this issue, referred to as Paper III.

2. Types of Backgrounds

There are at least three different types of background to which consideration must be given in a comprehensive study of detectors.

Radiation Background.—Every detector must operate in the presence of the blackbody radiation field appropriate to its own temperature and to the temperature of its environment. The noise due to this background has been termed *temperature noise*, and is discussed in detail in the first paper of this series.† This background is particularly important for detectors, such as thermocouples and bolometers, which are responsive in the 5 to 15 micron range.

Internal Background.—Another type of background is that introduced within the detector itself. Examples of such backgrounds are thermionic emission (dark current) in photoelectric cells and photo-multiplier tubes, Johnson noise, current noise,[1] and perhaps semi-conductor noise[2] in bolometers and photo-conductive cells, Johnson noise in thermocouples, and density fluctuations in photographic plates.

Signal Background.—If one is concerned with detecting a small change in a steady signal, the statistical fluctuations in the signal itself may determine the smallest change which may be detected. An example of this type of background is the increased noise in the output of a photo-tube when daylight is incident upon the tube. The effect of this background on the absolute and relative performance of the human eye, television systems, and the photographic plate has been discussed comprehensively by Rose.[3]

3. Types of Time Constants

The purpose of this Part is to establish that many of the known radiation detectors as they are customarily operated may be classified usefully on the way that their noise equivalent power** depends on their time constant and sensitive area, when one takes into account the limitations introduced by the radiation and internal backgrounds.

Before defining the classification, however, it is necessary to define the concept of the time constant. There are three different types of time constants used in the presentation.

There is first the *reference* time constant, defined in Section 7 on the basis of the response *versus* frequency curve obtained when the output has been equalized so that the noise power per unit band width is constant. This is probably the most fundamental definition of the time constant of a detector.

In general, any detector whose output is electrical may have its frequency response curve adjusted by electrical equalization so that the frequency response curve corresponds to any time constant. Any time constant obtained by using equalization different from that used in defining the reference time constant is termed the *effective* time constant.

Lastly, some detectors such as the bolometer possess one or more *physical* time constants. Examples of physical time constants are RC time constants, and the time constant obtained by taking the ratio of a heat capacity to a thermal conductivity coefficient.

In some cases, as for example the unequalized vacuum bolometer without current noise, these three differently defined time constants are all equal, but in other cases, they may be quite different.

4. Definitions Relating to the Responsivity

First, let the *responsivity* S of a detector with an electrical output be defined as the ratio of the output voltage to the power incident upon the detector. The responsivity S is conveniently expressed in volts per watt. The responsivity is a function of the modulation frequency of the radiation incident upon the detector, and in order to insure that the low frequency responsivity approaches the zero-frequency responsivity as a limit, it is necessary to postulate that the same measure of the amplitude be used for the radiation and for the voltage: for example, if the voltage is stated in terms of its root-mean-square value, then the incident power should be measured in terms of the root-mean-square value of the deviation of the incident power from its mean value. One may state, for example, that the responsivity is in r.m.s. volts per r.m.s. watt.

Secondly, let the *noise power* be defined as the mean square value of the noise voltage. Because the noise powers, rather than noise voltages, of non-overlapping frequency bands combine additively, it is convenient to deal with the noise power per unit band width, denoted by N.

Finally, the responsivity-to-noise ratio $\mathfrak{R}$ is defined as the ratio of the responsivity S to the square root of N:

$$\mathfrak{R} \equiv S/N^{\frac{1}{2}}. \tag{4.1}$$

† Attention is called to the erratum on page 343 with reference to Paper I.

[1] C. J. Christensen and G. L. Pearson, Bell Sys. Tech. J. **15**, 197–223 (1936).

[2] B. Davydov and B. Gurevich, "Voltage fluctuations in semi-conductors," J. Phys., USSR, **7**, 138–140 (1943). In Russian.

[3] Albert Rose, J. Opt. Soc. Am. **38**, 196–208 (1948); J. Soc. Mot. Pict. Eng. **47**, 273–294 (1946).

** The quantity which is here termed the noise equivalent power was denoted by the phrase minimum detectable power in Paper I. This change is made in order that the phrase used should not imply more than is involved in its definition.

5. The Sensitive Area

The sensitive area A of a detector is defined as the area over which the detector responds to irradiation. It will be supposed in this section that the detector is equally responsive at every part of its sensitive area.

It will now be shown, under certain restrictive conditions, that the responsivity-to-noise ratio $\mathfrak{R}$ is inversely proportional to the square root of the sensitive area.

The restrictive assumptions are:

1. It is assumed that the output voltage of the detector is proportional to the radiant power incident upon the sensitive area, when the power has a constant spectral distribution.
2. It is assumed that the noise contributed by a given part of the sensitive area is statistically independent of the noise contributed by any other non-overlapping part of the sensitive area.

Now consider two separate cases: (a) in which a certain power P is incident upon a single detector, and (b) in which the same total power is incident upon n adjacent detectors, each of which is identical with the detector in case (a). The total sensitive area in case (b) is thus n times that in case (a). Suppose further that the detectors in case (b) are connected in series-aiding.

Then according to the first assumption, the output voltage will be the same in the two cases, since in case (b) each of the detectors will be irradiated by the power P/n. On the other hand, by the second assumption, the noise voltage in case (b) will be greater than that in case (a) by the factor $n^{\frac{1}{2}}$. It now follows directly that the responsivity-to-noise ratio of any detector satisfying the assumptions is inversely proportional to the square root of the sensitive area.

The same conclusion follows from any method of connecting together the detectors which permits each of the n units to contribute equally to the total output in case (b).

Accordingly, in order to intercompare detectors with different sensitive areas, the responsivity-to-noise ratios should be multiplied by a factor proportional to the square root of the sensitive area of each detector.

This result is of rather general applicability, and applies also to at least one detector which is not associated with an electrical amplifier: the photographic plate.

The proof is not applicable, however, to the dipole antenna, because the voltage output is proportional to the square root of the power falling upon it; thus the first assumption is not valid. The second assumption is also invalid, but it is substantially correct when the adjacent dipoles are separated by not less than one-half wave-length; this is usually the case with arrays of dipoles.

The result of the above proof may be applied to dipole antennas, however, if each of the individual dipoles is separately connected to a square-law rectifier, and if the outputs of the separate rectifiers are connected in series or parallel. It must further be assumed that the dipoles are separated by at least one-half wave-length, in order that the noise outputs be substantially independent in the statistical sense. See Part II, Section 7.

6. Reference Condition of Measurement

In order to permit a straightforward intercomparison of detectors with different time constants, it is desirable that their noise equivalent powers be stated for a suitable standard condition of measurement.

A reference condition of measurement was proposed in the first paper of the series. A reformulation of this proposal is given in this section.

The reference condition of measurement described below satisfies the following conditions:

1. The noise equivalent power of the detector is measured in the presence of the noise in a manner such that the band width of the noise is approximately equal to the band width of the detector.
2. The band width of the detector is measured after the amplifier gain has been equalized so that the noise spectrum is flat.

Let the detector be coupled to an amplifier, and let it be required that the amplifier be such and the method of coupling be such that the noise power introduced by the amplifier is small at every frequency compared with the noise power in the output of the detector.

The frequency response curve of the amplifier is now to be adjusted so that the noise power per unit band width is the same at all frequencies.[4] With this type of equalization the frequency response curve of the detector is measured, and the time constant of the detector is determined from the frequency response curve in a manner discussed in detail in the following section. The time constant so obtained is called the *reference time constant.*

After the reference time constant has been determined, the frequency response curve of the amplifier is changed by adding to the circuits of the amplifier a high frequency RC cut-off (series resistance, shunt capacity) with a time constant equal to the reference time constant of the detector.

One then determines the steady input power to the detector which produces a steady voltage*** at

[4] In order that this specification be unique, it is necessary to require that the complex gain *versus* frequency function of the amplifier be of the minimum-theta type. See Hendrik W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, Inc., New York, 1945).

*** There is no difficulty of principle in measuring a steady response equal to the r.m.s. noise voltage, since an arbitrarily long time is available for measuring the zero-frequency response before and after the incident power is initiated.

the output of the amplifier which is equal to the root-mean-square noise voltage at the output of the amplifier. The power so measured is termed the *noise equivalent power of the detector in the reference condition A* and is denoted by P_m.

A relation between the zero-frequency responsivity-to-noise ratio $\mathfrak{R}_0$ and the noise equivalent power P_m will now be derived. Evidently, the noise equivalent power is the ratio of the r.m.s. noise voltage E_N to the zero-frequency responsivity S_0 of the detector

$$P_m = E_N/S_0. \tag{6.1}$$

If N_0 is the constant noise power per unit band width prevailing during the measurement of the reference time constant, then the noise power per unit band width during the measurement of the noise equivalent power is

$$N(f) = \frac{N_0}{1+(2\pi f\tau)^2}, \tag{6.2}$$

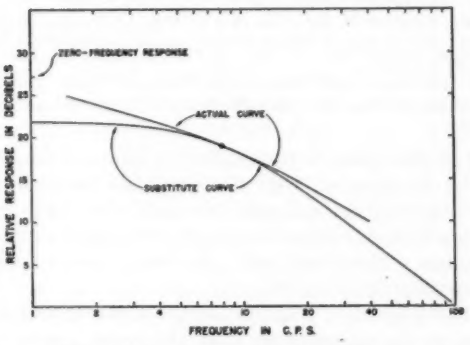

FIG. 1. Responsivity as a function of the frequency for the Schwarz thermopile, as measured by Nielsen. Since the noise is assumed to be Johnson noise only, the curve representing the reponsivity-to-noise ratio has the same shape. The curve slopes off more gradually at high frequencies than it does in the case of a detector with a single time constant. This characteristic defines a curve as a Class *I* curve.

the r.m.s. noise voltage E_N is then given by

$$E_N^2 = \int_0^\infty N(f)df = N_0/4\tau. \tag{6.3}$$

By the definition of $\mathfrak{R}$, one has

$$S_0 = N_0^{\frac{1}{2}}\mathfrak{R}_0. \tag{6.4}$$

Upon substituting Eqs. (6.4) and (6.3) in (6.1), one finds

$$P_m = 1/(2\tau^{\frac{1}{2}}\mathfrak{R}_0). \tag{6.5}$$

This is the desired relation.

Since it was established in Section 5 that the zero-frequency responsivity-to-noise ratio is inversely proportional to the square root of the sensitive area, it follows from the last equation that the noise equivalent power in reference condition *A* is proportional to the square root of the sensitive area. For this reason, it is convenient to define the *noise equivalent power in reference condition B*, denoted by H_m:

$$H_m \equiv P_m/A^{\frac{1}{2}}, \tag{6.6}$$

where A is here the sensitive area expressed in square millimeters. As used in this equation, the area A is dimensionless, and might more accurately be defined as the ratio of the sensitive area to an area of one square millimeter.

The noise equivalent power in reference condition *B* is the noise equivalent power the detector would have if its sensitive area were changed to one square millimeter.

The measurement of the noise equivalent power in the reference condition involves using a band width of noise equal to $\pi/2\tau$ in radians per second, or to $1/4\tau$ in cycles per second. In practice, it will not usually be convenient to measure the noise equivalent power with exactly this band-width of noise. If, however, the noise is known to be flat, then a measurement of the noise equivalent power with any other band width may be used to calculate the noise equivalent power in the reference condition. Furthermore, it will not usually be convenient in practice to measure the noise equivalent power for a zero-frequency signal. If, however, the $\mathfrak{R}$ *versus* frequency curve of the detector is known, then again one may calculate the noise equivalent power in the reference condition. For the sake of specificity, the method of calculation is stated in the following paragraph.

Let the noise equivalent power be measured with an amplifier whose gain curve is such that the power spectrum of the noise is $N(f)$, and with a sinusoidal radiation signal of frequency f_s; let the noise equivalent power so obtained be denoted by P. Let the noise band width Δf of the amplifier be defined by

$$\Delta f = \frac{\int_0^\infty N(f)df}{N_{\max}}. \tag{6.7}$$

Furthermore, let the zero-frequency value of $\mathfrak{R}$ be denoted as before by $\mathfrak{R}_0$, and let the value of $\mathfrak{R}$ at the signal frequency be denoted by $\mathfrak{R}_s$. Then the noise equivalent power in the reference condition *A* is obtained from P by

$$P_m = P\cdot\frac{\mathfrak{R}_s}{\mathfrak{R}_0}\cdot\left(\frac{1/4\tau}{\Delta f}\right)^{\frac{1}{2}}\cdot\left(\frac{N(f_s)}{N_{\max}}\right)^{\frac{1}{2}}, \tag{6.8}$$

where τ is the reference time constant.

A somewhat different definition of the noise equivalent power in the reference condition was given in the first paper of this series. The definition given there involved the same specifications as those used here with regard to the noise spectra, but involved the use of a sinusoidal signal with the frequency $1/4\tau$ instead of the steady signal specified here. For detectors in accord with Eq. (7.1), the present noise equivalent power is one-half of that defined in Paper *I*. The extension of the definition of the noise equivalent power to detectors not in accord with Eq. (7.1), however, is accomplished

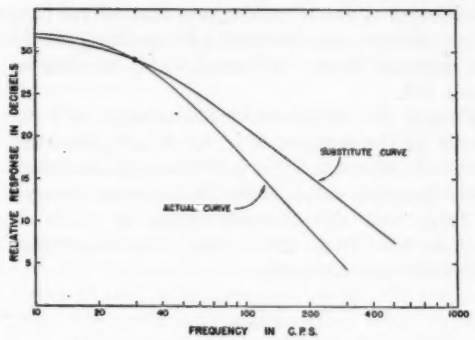

FIG. 2. Responsivity as a function of the frequency for the Strong bolometer, as measured by Nielsen. Since the noise is assumed to be Johnson noise only, the curve representing the responsivity-to-noise ratio has the same shape. The curve falls off more rapidly at high frequencies than it does in the case of a single time constant. This characteristic defines the curve as a Class *II* curve.

much more naturally with the present definition, and it is for this reason that the present definition was introduced.

Lest the reader receive from this section the impression that the noise equivalent power defined here is equal to the minimum detectable power under the same conditions, it should be pointed out that Rose[3] has found that in certain visual phenomena the minimum detectable stimulus is about five times the noise equivalent stimulus.

7. Determination of the Reference Time Constant

As specified in the last section, the reference time constant is to be determined from the shape of the responsivity-to-noise ratio *versus* frequency curve.

Consider first the case in which the shape of this curve corresponds to the existence of a single time constant. By this one means that the responsivity-to-noise ratio depends upon the frequency according to

$$\mathfrak{R}(f)=\frac{\mathfrak{R}_0}{(1+(2\pi f\tau)^2)^{\frac{1}{2}}}, \qquad (7.1)$$

where τ is the single time constant. This case offers no difficulty; the reference time constant is set equal to the time constant τ.

When $\mathfrak{R}$ does not vary with frequency in accordance with Eq. (7.1), it is frequently stated that the detector "does not have a time constant," or that it "has more than one time constant." The latter method of expression has the following significance. If $\mathfrak{R}$ is expressed as a complex function that indicates both the amplitude and the phase of the response of the detector, then by a general mathematical theorem[5] relating to analytic functions without essential singularities, the function $\mathfrak{R}(f)$ can be completely specified in terms of its poles and zeros. Equation (7.1) corresponds to the existence of one zero at infinity and to one pole on the imaginary frequency axis. If further zeros and/or poles are involved, the dependence of $\mathfrak{R}$ on the frequency is more complicated.

Even in the case of complicated dependences of $\mathfrak{R}$ on the frequency, however, it is possible and convenient to define a single time constant for the detector. Perhaps the most simple, and at the same time the most generally useful definition of the reference time constant is

$$\tau=\frac{\frac{1}{4}\mathfrak{R}_{\max}{}^2}{\int_0^\infty \mathfrak{R}^2 df}, \qquad (7.2)$$

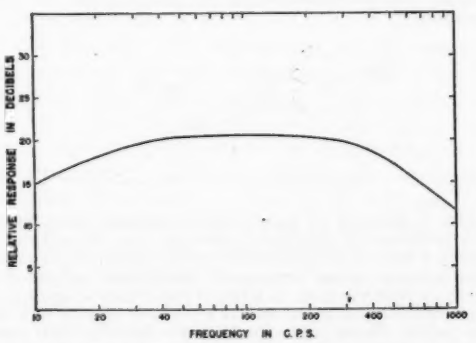

FIG. 3. Responsivity as a function of the frequency for the Golay pneumatic heat detector, as measured by Nielsen. Since the noise is assumed to be Johnson noise only, the curve representing the responsivity-to-noise ratio has the same shape. The defining characteristic of the Class *III* type of curve is that the responsivity-to-noise ratio falls off at low frequencies.

where $\mathfrak{R}_{\max}$ is the maximum value of $\mathfrak{R}$, and where the numerical constant $\frac{1}{4}$ has been introduced in order that Eq. (7.2) should agree with Eq. (7.1).

In all of the specific detectors actually studied in the fifth paper of this series, however, the nature

[5] Liouville's Theorem. See E. T. Whittaker and G. N. Watson, *Modern Analysis*, fourth edition (Cambridge University Press, London, 1927), p. 105.

of the $\mathfrak{R}$ *versus* frequency curves is such that substantially the same result as that obtained from Eq. (7.2) can be obtained by procedures which substitute a curve representing a single time constant for the actual curve. The reference time constant and the effective value of the zero-frequency responsivity-to-noise ratio are both defined in terms of the substitute curve.

Three different classes of responsivity-to-noise ratio *versus* frequency curves were encountered which did not correspond to the existence of a single time constant. They are described as follows:

Class I

The most common type of departure from Eq. (7.1) is that illustrated in Fig. 1. This figure shows

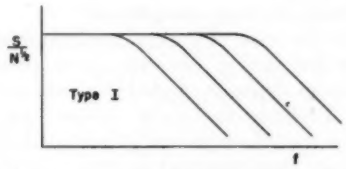

FIG. 4. Showing on logarithmic coordinate scales the relation between the responsivity-to-noise ratio $\mathfrak{R}$ and the frequency f for a Type I detector with a single time constant. The separate curves correspond to different values of the reference time constant. The curves all approach asymptotically the same horizontal line at low frequencies, in accordance with the fact that the zero-frequency responsivity-to-noise ratio is independent of the time constant for Type I detectors.

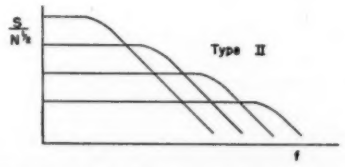

FIG. 5. Showing on logarithmic coordinate scales the relation between the responsivity-to-noise ratio $\mathfrak{R}$ and the frequency f for a Type II detector with a single time constant. The separate curves correspond to different values of the reference time constant. A straight line drawn tangent to the knees of the separate curves has a negative slope of 3 decibels per octave. These curves make very clear the fact that in order to maximize the responsivity-to-noise ratio for a sinusoidal signal with angular frequency ω, the time constant of a Type II detector should be set equal to $1/\omega$.

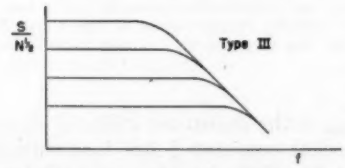

FIG. 6. Showing on logarithmic coordinate scales the relation betweer. the responsivity-to-noise ratio $\mathfrak{R}$ and the frequency f for a Type I detector with a single time constant. The separate curves correspond to different values of the reference time constant. The curves all approach asymptotically the same straight line at high frequencies. This line has a negative slope of 6 decibels per octave.

the $\mathfrak{R}$ *versus* frequency curve measured by Nielsen[6] for the Schwarz No. 11 thermopile. The defining characteristic of this curve is that it slopes off more gradually as the frequency increases than it does in the case of a single time constant, even though the asymptotic slope may be as large as or even greater than 6 db per octave. For a Class I response curve it is here proposed that the actual curve be replaced by a substitute curve characterized by a single time constant in such a way that the substitute curve and the actual curve have in common the points at which they individually have slopes of 3 db per octave. Of all of the possible ways of making the two curves tangent to one another, the method just specified yields the the greatest factor of merit, M_2, as defined in Paper III.

Thus if the actual curve has a slope of 3 db per octave at the frequency f_0, at which point the responsivity-to-noise ratio is N decibels less than its zero-frequency value, then the reference constant is $1/(2\pi f_0)$, and the effective value of $\mathfrak{R}_0$ is $N-3$ decibels less than the actual zero-frequency responsivity-to-noise ratio.

Class II

This type of curve is illustrated in Fig. 2. This figure shows the curve measured by Nielsen for the Strong bolometer. The defining characteristic of this class of curves is that at high frequencies the curve falls off more rapidly than is the case with a single time constant. For such detectors the reference time constant may usefully be defined in terms of the frequency at which $\mathfrak{R}$ is 3 db less than its value at zero frequency. The effective zero-frequency responsivity-to-noise ratio is then to be taken equal to the actual value of $\mathfrak{R}_0$.

Class III

Class III is illustrated by the curve shown in Fig. 3. This figure represents the responsivity-to-noise ratio as measured by Nielsen for the Golay pneumatic heat detector. The defining characteristic of the class is that the behavior at high frequencies may either be that corresponding to Eq. (7.1) or be that corresponding to the behavior represented by a Class I or II detector, but that the curve falls off at low frequencies. In this case, provided the flat portion of the curve is several octaves wide, it is convenient and useful to extend the flat portion of the curve to zero frequency and to define the reference time constant and the effective zero-frequency responsivity-to-noise ratio in terms of the substitute curve so obtained.

[6] Harald H. Nielsen, *Comparative Testing of Thermal Detectors*, OSRD Report No. 5992, Contract OEMsr-1168 (October 31, 1945).

8. Detectors with Non-Uniform Spectral Response

Up to this point the nature of the radiation in terms of which the responsivity is to be measured has been left undefined. If the response is not the same for all wave-lengths of the radiation signal, then it is clearly necessary to introduce a parameter upon which the noise equivalent power must depend. In many cases it will be most convenient to take this parameter to be the wave-length of the signal, that is to say, the noise equivalent power must be expressed as a function of the wave-length of the radiation. If the factor of merit defined in Paper III is expressed in these terms, then the factor of merit for any given spectrum may be determined by averaging the factor of merit with respect to the spectrum.

In other cases, however, it will be more convenient to use the effective temperature of the radiation as the parameter. In this case the noise equivalent power and factor of merit are expressed as functions of the effective temperature of the radiation. This method of description has the disadvantage that when the absolute radiation temperature is not large compared with the temperature of the detector, the response then depends upon both of these temperatures. This difficulty may be avoided in part by stating the response the detector would have to radiation of a given absolute temperature if the detector were at the temperature absolute zero but had otherwise the same properties.

9. The Proposed Classification

On the basis of the foundation established in the preceding sections of this paper, it is now possible to present a simple statement of the proposed classification system for radiation detectors.

A detector is here defined to be a Type n (Type I, Type II, etc.) detector for a given range of reference time constants if over that range of reference time constants the zero-frequency responsivity-to-noise ratio $\mathcal{R}_0$ depends upon the area A and the reference time constant τ in accordance with

$$\mathcal{R}_0=\frac{k_n\tau^{\frac{1}{4}(n-1)}}{2A^{\frac{1}{4}}}, \tag{9.1}$$

where k_n is a constant which is independent of A and τ.

In view of the relation (6.5), this definition is equivalent to the statement that a detector is a Type n detector if over the given range of reference time constants the noise equivalent power in the reference condition A is related to the sensitive area A and the time constant τ according to

$$P_m=(A^{\frac{1}{4}}/k_n\tau^{\frac{1}{4}n}). \tag{9.2}$$

Either Eq. (9.1) or (9.2) may be considered as the definition of a Type n detector.

By combining Eqs. (7.1) and (9.1) one finds

$$\mathcal{R}(f)=\frac{k_n\tau^{\frac{1}{4}(n-1)}}{2A^{\frac{1}{4}}(1+(2\pi f\tau)^2)^{\frac{1}{4}}}. \tag{9.3}$$

The relation defined by this equation is exhibited graphically in Figs. 4 to 6, with regard to the dependence of $\mathcal{R}$ on f and τ. In each of these figures the separate curves correspond to different fixed values of the time constant. The figures utilize logarithmic coordinate scales.

In the case of the Type I detector shown in Fig. 4, the curves representing different time constants all approach asymptotically the same horizontal line, and in the case of the Type III detector shown in Fig. 6, the curves representing different time constants all approach asymptotically the same sloping line. In the intermediate case of the Type II detector, shown in Fig. 5, every curve crosses every other curve. Once one has clearly understood the significance of the curves in these figures, they represent the easiest way of remembering the distinction between the different types of detector behavior.

All of the detectors studied in Part II are either Type I or Type II detectors.

A list of the classification of various specific types of detectors is given in Table I of Part II.

With regard to the dependence on the reference time constant, the zero-frequency responsivity-to-noise ratio of a Type I detector is independent of the time constant, whereas with a Type II detector, $\mathcal{R}_0$ is proportional to the square root of the time constant. Similarly, the noise equivalent power in the reference condition of a Type I detector is inversely proportional to the square root of the time constant, whereas with a Type II detector, P_m is inversely proportional to the time constant.

In considering the use of any given type of a detector for a specific application, it is important to know the type number of the detector because the optimum proportioning of the equipment with which the detector is associated depends on the type number of the detector. This situation is exemplified in the concluding section of this Part.

10. An Application to Sequential Scanning Systems

In this section a brief application of the classification system established in the preceding section will be presented. The chief purpose of this section is to illustrate the usefulness of the classification, but the problem studied is also of considerable interest in its own right.

The problem set is as follows. Consider a sequential scanning system whose instantaneous field of view covers the solid angle ω and which scans at such a rate that a total solid angle $\dot{\Omega}$ is scanned in unit time. The field to be scanned contains no sources of a radiation of the type to be detected, except for one or two point targets. The only sources of noise are thus the thermodynamic and the internal backgrounds. The problem is to calculate the power of the target which yields a detector response equal to the r.m.s. noise of the detector.

The detecting system will be assumed to consist of a lens which focuses upon the detector the field of view at infinity. The detector responds to the total power incident upon its sensitive area. The scanning process will consist simply of a movement of the entire optical system in such a way that it views successively the various parts of the total field of view which is to be scanned.

The limitation on the sources of noise means that the present treatment will not apply, for example, to television systems as they are ordinarily employed. The limiting background in ordinary television systems is the third type of background discussed in Section 2.

The Stationary System

An expression will be derived in this section for the target strength which produces a steady signal voltage which is equal to the r.m.s. noise voltage, when the optical system is trained steadily in a given direction.

Let S be the strength of the target, as measured by the radiant flux which it produces at the location of the detecting system, in watts/cm^2. Then if the target is a point source, the power collected by the lens and thereby made to fall on the detector is equal to the product of the target strength S, the effective area of the collecting lens, and the energy transmission factor T:

$$P=\tfrac{1}{4}\pi D^2TS, \tag{10.1}$$

where D is the effective diameter of the collecting lens.

The area of the sky which subtends the solid angle ω and which is focused upon the sensitive area A of the detector by the lens is defined by

$$\omega=A/f^2, \tag{10.2}$$

where f is the focal length of the lens. The numerical aperture N of the lens is defined by

$$N\equiv D/2f. \tag{10.3}$$

By combining these three relations, one finds

$$P=\tfrac{1}{2}\pi NDT(A/\omega)^{\frac{1}{2}}S. \tag{10.4}$$

If now in this expression the power P is set equal to the noise equivalent power P_m of the detector, then the target strength S becomes equal to the noise equivalent target strength S_m:

$$P_m=\tfrac{1}{2}\pi NDT(A/\omega)^{\frac{1}{2}}S_m. \tag{10.5}$$

According to the classification system developed in Section 9, the noise equivalent power P_m of a Type n detector depends upon its area A and its reference time constant τ according to

$$P_m=(A/\tau^n)^{\frac{1}{2}}/k_n, \tag{10.6}$$

where k_n is a constant independent of A and τ.

Upon eliminating P_m between the last two equations, and solving for S_m, one finds

$$1/S_m=\tfrac{1}{2}\pi k_nNDT(\tau^n/\omega)^{\frac{1}{2}}. \tag{10.7}$$

The discussion of the significance of this expression will be deferred until it has been modified by introducing the scanning process.

The Scanning Process

It will now be assumed that the optical system is moved continuously so that it scans over solid angle at the rate $\dot{\Omega}$.

It will further be assumed that the rate of scanning is such that each element of solid angle ω is surveyed for a length of time equal to the reference time constant τ. One then has

$$\dot{\Omega}=\omega/\tau. \tag{10.8}$$

The expression (10.7) gives the noise equivalent target strength for a steady signal. If the signal has a duration of only τ, the noise equivalent target strength will be approximately doubled.**** If this correction is taken into account, one finds from Eqs. (10.7) and (10.8)

$$1/S_m=\tfrac{1}{4}\pi k_nNDT(\omega^{n-1}/\dot{\Omega}^n)^{\frac{1}{2}}. \tag{10.9}$$

Discussion

All of the results of the present treatment are contained in the last equation.

Independent of the type of the detector, the sensitivity of the system, as measured by $1/S_m$, is

**** The factor is exactly two for a sinusoidal signal with the frequency $f=1/2\pi\tau$. The signal voltage is decreased 3 decibels below its zero frequency value because of the decreased response of the detector itself, and 3 decibels more because of the decreased gain of the amplifier in the reference condition.

directly proportional to the constant k_n of the detector, to the numerical aperture N, to the linear size of the collecting lens, and to the over-all transmission of the optical system.

The last factor

$$(\omega^{n-1}/\dot{\Omega}^n)^{\frac{1}{2}} \tag{10.10}$$

of Eq. (10.9) is the only one which depends on the type number of the detector. If the detector is of Type I, this factor is simply

$$1/\dot{\Omega}^{\frac{1}{2}}. \tag{10.11}$$

Thus one finds the important result that with a Type I detector the noise equivalent target strength is independent of the resolution $1/\omega$ of the scanning system, and depends only on the rate at which solid angle is scanned.

For a Type II detector, the situation is quite different. In this case, the factor (10.10) becomes

$$\omega^{\frac{1}{2}}/\dot{\Omega}. \tag{10.12}$$

It follows that with a Type II detector the signal-to-noise ratio is adversely affected by an increase in resolution. It further follows that a heavier price must be paid for increasing the scanning rate $\dot{\Omega}$ than is paid with a Type I detector.

PART II. CLASSIFICATION OF EIGHT SPECIFIC KINDS OF DETECTORS

1. Introduction

This Part presents the considerations on the basis of which the type numbers of eight different kinds of radiation detectors have been assigned. The system of classification is that proposed in Part I, Section 9.

The conclusions of this Part are summarized in Table I. This table lists the Type I and Type II detectors separately, and states the range of reference time constants for which each classification holds.

For those kinds of detectors whose limiting noise is Johnson noise, the relation between the zero-frequency responsivity $\mathcal{S}_0$ and the zero-frequency responsivity-to-noise ratio $\mathcal{R}_0$ will be derived in this section for later use.

The relation will be derived on the basis of the assumption that the detector works into an open circuit. In this case, the noise power per unit frequency band width is given by

$$N_0 = 4kTR, \tag{1.1}$$

where k is Boltzmann's constant, T is the absolute temperature, and R is the electrical resistance of the detector. By the use of Eq. (4.1) of Part I, one then obtains

$$\mathcal{R}_0 = \mathcal{S}_0/(4kTR)^{\frac{1}{2}}. \tag{1.2}$$

This is the desired relation for detectors whose sensitivity is limited by Johnson noise.

TABLE I. Summary of the classification of the eight kinds of detectors studied in Part II. The range of attainable reference time constants for each of the detectors is also shown, as well as the section in which the detector is studied.

Detector	Range of τ in seconds	Section number
Type I detectors: $P_m \sim A^{\frac{1}{2}}/\tau^{\frac{1}{2}}$.		
Golay pneumatic heat detector	?	4
Vacuum phototubes limited by shot noise	$10^{-9} - \infty$	6
Gas photo-tubes limited by shot noise	$10^{-3} - \infty$	6
Photo-multiplier tubes limited by shot noise	$2.5 \times 10^{-10} - \infty$	6
Dipole antenna limited by temperature noise = Johnson noise	$10^{-9} - \infty$	7
Type II detectors: $P_m \sim A^{\frac{1}{2}}/\tau$		
Bolometers	$10^{-3} - 1$	2
Thermocouples and thermopiles	$10^{-2} - 1$	3
Photographic plates limited by grain structure	Reciprocity law	5

2. Bolometers

The discussion of bolometers is presented stepwise. It is first assumed that the only noise present is Johnson noise (internal background). It is then assumed that the limiting noise is temperature noise (radiation background). Finally, an expression is obtained which takes into account simultaneously both types of noise.

The zero-frequency responsivity of a bolometer may be written

$$\mathcal{S}_0 = \alpha R I \Delta T / P, \tag{2.1}$$

where α is the fractional change in the resistance per degree of temperature change, R is the resistance, I is the current through the bolometer, and ΔT is the rise in temperature of the bolometer when the radiant power P is incident upon its sensitive area. The change in temperature ΔT may be written

$$\Delta T = \epsilon P / \kappa A, \tag{2.2}$$

where ϵ is the emissivity of the bolometer, κ is the steady-state heat flow from the bolometer to its surroundings per unit area per degree of temperature difference, and A is the sensitive area.

If C is the heat capacity of the bolometer per unit area, the physical time constant is

$$\tau = C/\kappa, \tag{2.3}$$

and upon eliminating ΔT and κ by Eqs. (2.2) and (2.3), Eq. (2.1) becomes:

$$\mathcal{S}_0 = \epsilon \alpha R I \tau / CA. \tag{2.4}$$

Recognition will now be given to the fact that the permissible current I is usually limited by the fact that the current heats the bolometer, and that the maximum permissible temperature rise is limited. Let θ be the maximum permissible temperature rise. Then the maximum current is related to θ by

$$I^2R=\theta\kappa A=\theta CA/\tau. \tag{2.5}$$

Upon eliminating I between Eqs. (2.4) and (2.5), one finds

$$S_0=\epsilon\alpha(\theta R\tau/CA)^{\frac{1}{2}} \tag{2.6}$$

If now the sensitivity of the bolometer is limited by the internal background of Johnson noise, the physical time constant is equal to the reference time constant, and Eq. (1.2) may be employed, with the result

$$\mathcal{R}_0=\frac{\epsilon\alpha}{2}\cdot\left(\frac{\theta}{kTC}\right)^{\frac{1}{2}}\cdot\frac{\tau^{\frac{1}{2}}}{A^{\frac{1}{2}}}. \tag{2.7}$$

By comparison of this relation with Eq. (9.1) of Part I one concludes correctly that the bolometer is a Type *II* detector.

This conclusion is as yet upon a somewhat shaky basis, however, because if C is eliminated between (2.3) and (2.7), one obtains

$$\mathcal{R}_0=\frac{\epsilon\alpha}{2}\cdot\left(\frac{\theta}{kT\kappa}\right)^{\frac{1}{2}}\cdot\frac{1}{A^{\frac{1}{2}}} \tag{2.8}$$

from which one might conclude that the bolometer were a Type *I* detector.

In order to resolve this situation, one must introduce the following practical consideration: In constructing a bolometer one first makes C as small as possible, because this shortens the time constant without affecting the zero-frequency responsivity. After this is done, κ is increased by various means until the time constant has been shortened to the desired value. This description is considerably oversimplified, but is substantially correct. The chief respect in which this description is incomplete is that most of the methods ordinarily used to increase the value of the constant κ do so in a way which leads to the introduction of further time constants; in other words, the quantity κ does not correspond to a simple resistance but to a two-terminal network containing reactive elements as well as resistances.

Since C is a constant which for a given type of bolometer is independent of the time constant, it follows from Eq. (2.7) that the bolometer limited by Johnson noise is a Type *II* detector for the range of reference time constants over which the time constant is adjustable by varying the cooling constant κ. By inspection of Figs. 1 and 2 of Paper III, this range is seen to be roughly from one millisecond to one second.

It is possible to make bolometers with reference time constants substantially longer than one second by increasing the value of the thermal capacity per unit area C. As indicated by Eq. (2.8), the bolometer is then a Type *I* detector. No advantage is gained from this procedure so that it is not used in practice. On the other hand, it was just this type of variation which was used in obtaining the maximum possible sensitivity of a bolometer or thermocouple in Paper I. In obtaining this maximum possible sensitivity it was assumed that the cooling constant κ had its minimum possible value, namely the value obtained when the only cooling is that due to radiation. Under these conditions the only way of varying the time constant is to vary the thermal capacity C. Although it is recognized that in each particular design of bolometer or thermocouple the minimum attainable value of C will always be used, it is nevertheless true that in different designs of bolometers or thermocouples the minimum attainable value of C will be different.

From Eq. (2.8) it may be found by the use of Eq. (6.5) of Part I that the noise equivalent power P_m is given by

$$P_m=\frac{(kT^2CA)^{\frac{1}{2}}}{\epsilon\tau}\cdot\left(\frac{1}{\alpha^2\theta T}\right)^{\frac{1}{2}}. \tag{2.9}$$

So far in this section it has been assumed that the only noise present is the Johnson noise associated with the electrical resistance of the bolometer. The noise equivalent power in the reference condition A will now be calculated by assuming that the only noise present is temperature noise.††

According to Eq. (2.4) of Paper I the r.m.s. temperature fluctuation $\langle\Delta T\rangle_{\text{r.m.s.}}$ of the bolometer due to temperature noise is given by

$$\langle\Delta T\rangle_{\text{r.m.s.}}=(kT^2/CA)^{\frac{1}{2}}. \tag{2.10}$$

When this relation is combined with Eqs. (2.2) and (2.3) one finds

$$P_m=\frac{(kT^2CA)^{\frac{1}{2}}}{\epsilon\tau}. \tag{2.11}$$

This equation expresses the noise equivalent power in reference condition A if the only source of noise is temperature noise.

Consider now the actual case in which both types of noise are simultaneously present. First let it be noted that the noise powers of the two noises will

†† The properties of temperature noise are discussed at length in Paper I of this series.

simply add because the two noise voltages are statistically independent. Since the noise equivalent power P_m is proportional to the r.m.s. noise voltage, it then follows that the noise equivalent power in the presence of both types of noise is equal to the square root of the sum of the squares of the noise equivalent powers defined by Eqs. (2.9) and (2.11):

$$P_m=\frac{(kT^2CA)^{\frac{1}{2}}}{\epsilon\tau}\left(1+\frac{1}{\alpha^2\theta T}\right)^{\frac{1}{2}}. \qquad (2.12)$$

If one compares this relation with Eq. (9.2) of Part I, one finds that the bolometer, whether it is limited by Johnson noise, by temperature noise, or by both, is a Type *II* detector over the range of time constants attainable by varying the constant κ.

In Eq. (2.12) the first term in the last factor represents the effect of temperature noise and the second term represents the effect of Johnson noise.

3. Thermocouples and Thermopiles

The zero-frequency responsivity of a thermocouple or a series-connected thermopile may be written

$$S_0=nQ\Delta T/P, \qquad (3.1)$$

where n is the number of hot junctions, Q is the thermoelectric power of the pair of metals used, and ΔT is the change in the temperature of the hot junctions when the radiant power P is incident thereon. The change in temperature ΔT may be written

$$\Delta T=\frac{\epsilon P}{\kappa A+n(\sigma_1+\sigma_2)}, \qquad (3.2)$$

where ϵ is the emissivity, κ is the heat flow from the receiver attached to each of the hot junctions to the surroundings per unit area per unit temperature difference by all means except by the electrical conductors, σ is the heat flow per unit temperature difference from a single hot junction to a single cold junction through the electrical conductor, A is the total area of the receivers attached to the hot junctions, and the subscripts 1 and 2 refer to the two metals.

The physical time constant of the thermopile is then given by

$$\tau=\frac{CA}{\kappa A+n(\sigma_1+\sigma_2)}, \qquad (3.3)$$

where C is the heat capacity per unit area. The conductivities σ_1 and σ_2 may be written

$$\sigma_1=k_1a_1/l_1, \qquad (3.4)$$

$$\sigma_2=k_2a_2/l_2, \qquad (3.5)$$

where k_1 is the thermal conductivity of the electrical conductor of metal No. 1, a_1 is the effective cross-sectional area of this conductor, and l_1 is the length of this conductor. By combining Eqs. (3.2) and (3.3) one finds the following simple expression for the zero-frequency responsivity:

$$S_0=nQ\epsilon\tau/CA. \qquad (3.6)$$

It is now necessary to derive the relation between the quantity n in the last equation and the resistance R of the thermocouple. If the couples are connected in series, the resistance R of the thermopile may be written

$$R=n(l_1\rho_1/a_1+l_2\rho_2/a_2), \qquad (3.7)$$

where ρ is the electrical resistivity. By the use of Eqs. (3.4) and (3.5) the last relation may be written

$$R=n(k_1\rho_1/\sigma_1+k_2\rho_2/\sigma_2). \qquad (3.8)$$

The resistance R will now be minimized, subject to the condition that $\sigma_1+\sigma_2$ has a constant value. This procedure leads to the following results:

$$\frac{\sigma_1}{\sigma_2}=\left(\frac{k_1\rho_1}{k_2\rho_2}\right)^{\frac{1}{2}}, \qquad (3.9)$$

$$\frac{a_1l_2}{a_2l_1}=\left(\frac{k_2\rho_1}{k_1\rho_2}\right)^{\frac{1}{2}}. \qquad (3.10)$$

By the use of Eq. (3.9), the resistance R may be written

$$R=\frac{n}{\sigma_1+\sigma_2}((k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}})^2, \qquad (3.11)$$

and upon combining Eqs. (3.3) and (3.11), one finds

$$R=\frac{n^2\tau}{CA}((k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}})^2\left(1+\frac{\kappa A}{n(\sigma_1+\sigma_2)}\right). \qquad (3.12)$$

By solving the last relation for n as it appears in the first factor of the last relation and substituting the result in Eq. (3.6), one finds the following expression for the zero-frequency responsivity of the thermopile:

$$S_0=\frac{\epsilon Q}{(k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}}}\cdot\left(1+\frac{\kappa A}{n(\sigma_1+\sigma_2)}\right)^{-\frac{1}{2}}\left(\frac{R\tau}{CA}\right)^{\frac{1}{2}}. \qquad (3.13)$$

If now it be assumed that the only noise is that due to the internal background of Johnson noise associated with the resistance R, one finds by the use of Eq. (1.2) that the noise equivalent power is given by

$$\mathcal{R}_0=\frac{\epsilon Q}{(k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}}}\cdot\frac{1}{(kTC)^{\frac{1}{2}}}\cdot\left(1+\frac{\kappa A}{n(\sigma_1+\sigma_2)}\right)^{-\frac{1}{2}}\cdot\frac{\tau^{\frac{1}{2}}}{A^{\frac{1}{2}}}. \quad (3.14)$$

It follows from the last equation that in an efficient design the cooling of the hot junction by conduction through the wires is made large compared with the cooling by all other causes, that is to say, one satisfies the condition

$$\kappa A \ll n(\sigma_1+\sigma_2). \quad (3.15)$$

By making use of the same practical consideration invoked in Section 2, it now follows from the last two equations that the thermocouple limited by Johnson noise is a Type *II* detector. By inspection of Figs. 1 and 2 of Paper III, one finds that this conclusion holds for time constants lying between about 10 milliseconds and one second.

By use of Eq. (6.5) in Part I and Eq. (3.14) in this section, the noise equivalent power P_m is given by

$$P_m=\frac{(kT^2CA)^{\frac{1}{2}}}{\epsilon\tau}\cdot\frac{(k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}}}{2QT^{\frac{1}{2}}}\left(1+\frac{\kappa A}{n(\sigma_1+\sigma_2)}\right)^{\frac{1}{2}}. \quad (3.15)$$

So far in this section it has been assumed that the only noise present is Johnson noise. By use of the same method employed in Section 2 and in particular by the use of Eq. (2.11), the effect of temperature noise may be introduced into the result (3.15). One finds

$$P_m=\frac{(kT^2CA)^{\frac{1}{2}}}{\epsilon\tau}\cdot\left\{1+\frac{[(k_1\rho_1)^{\frac{1}{2}}+(k_2\rho_2)^{\frac{1}{2}}]^2}{4Q^2T}\left(1+\frac{\kappa A}{n(\sigma_1+\sigma_2)}\right)\right\}^{\frac{1}{2}}. \quad (3.16)$$

It follows from this equation that the thermocouple and thermopile, whether they are limited by Johnson noise, by temperature noise, or by both, are Type *II* detectors.

The conclusions of this section with regard to the classification of the thermopile are in no way changed if one assumes that the separate pairs of junctions are connected in parallel or in series-parallel.

4. The Golay Pneumatic Heat Detector

The theoretical basis of this detector has been described[7] at length by its inventor, and will not be repeated here.

The type number of this detector depends upon how one modifies the construction of the detector in order to change its time constant. The nature of the construction is such as to suggest that the constant κ is set in advance, and that the time constant can be adjusted only by changing the thermal capacity per unit area of the membrane. If this is true, the Golay pneumatic heat detector is a Type *I* detector, and it is provisionally so classified in Table I.

The theory indicates that the responsivity-to-noise ratio should be less by the factor $3^{\frac{1}{2}}$ than the theoretical maximum calculated in Paper I. The measurements reported in Paper III suggest an actual factor of 3.3.

5. The Photographic Plate

In this section the type number of the photographic plate will be determined on the basis of the assumption that the smallness of the minimum detectable energy is limited by the spatial fluctuations of the density within a uniformly exposed area (internal background).

Because the photographic plate does not yield an electrical voltage as its output, the reference condition of measurement defined in Part I does not apply to it. The chief problem accordingly lies in the definition of the reference time constant to be ascribed to the photographic plate. The procedure most closely analogous to that used with other detectors would be to define the time constant in terms of the characteristic times associated with the build-up and decay of the latent image. These time constants may be very long, of the order of days, months, or years.

But photographic plates are rarely used in a way such that the build-up and decay times of the latent image are the controlling ones. In detectors with an electrical output, the significance of the reference time constant τ is that if one uses the same equalization in the amplifier as that used in the reference condition of measurement, the noise equivalent power is increased only slightly if one

[7] Marcel J. E. Golay, Rev. Sci. Inst. 18, 347–362 (1947).

repeats the measurement every τ seconds. Accordingly, the writer postulates that the duration of the exposure of the plate is the reference time constant of the photographic plate, and that it is the time constant so defined that may most meaningfully be compared with the reference time constants of other detectors.

It is assumed that the plate has been given a uniform pre-exposure in order to bring the plate to the optimum point on its characteristic curve, and that the optimum method of development is employed. The optimum condition may be defined as that which maximizes the ratio k/κ, where k and κ are defined in Eqs. (5.1) and (5.3). Then, because of the distribution in the size of the developed grains, and because of their more or less random spatial distribution, there is a fluctuation of the density from one location on the plate to another. Furthermore, if one measures the density of a large number of different portions, each of area A, of the uniformly exposed region, there will be a fluctuation in the measured density. This fluctuation means that the average density of any given region of area A is not a unique function of the power incident upon the area, but contains also a random element. The fluctuation may be compared with the noise in electrical detectors, in that it serves to place a lower limit on the power which may be detected.

If $\langle \Delta D \rangle_{\text{r.m.s.}}$ is the r.m.s. fluctuation of the density in the area A, the fluctuation equivalent exposure of the plate is that exposure over the area A which produces a change in density equal to $\langle \Delta D \rangle_{\text{r.m.s.}}$.

Now consider the change in density ΔD produced on the plate by a certain power P incident upon the area A during the period τ. Because of the curvature of the characteristic curve of the photographic plate, the change in density ΔD will not be exactly proportional to the energy incident per unit area $P\tau/A$. Provided, however, that the fractional variation of the slope of the characteristic curve is small over the range of exposures considered, one has approximately

$$\Delta D = kP\tau/A, \tag{5.1}$$

where k is a constant of proportionality.

It then follows from the last relation that the fluctuation equivalent power is

$$P_m = A\langle \Delta D \rangle_{\text{r.m.s.}}/k\tau \tag{5.2}$$

subject to the restriction stated above.

The r.m.s. fluctuation of the density has been shown by a number of recent measurements[8] to be proportional to the inverse square root of the area A

$$\langle \Delta D \rangle_{\text{r.m.s.}} = \kappa A^{-\frac{1}{2}}, \tag{5.3}$$

where κ is another constant of proportionality. This relation is restricted by the condition that the area A must be so large that the average number of grains contained within it is large compared with unity. Equation (5.3) will usually hold[9] for areas not smaller than about 0.005 mm².

By combining the last two relations, one finds

$$P_m = (\kappa/k)(A^{\frac{1}{2}}/\tau). \tag{5.4}$$

This relation indicates that the photographic plate is a Type *II* detector, subject to the two restrictions involved in Eqs. (5.1) and (5.3). Both of the restrictions have the effect of limiting the smallness of the area A.

The result (5.4) is limited in a further way, however, through the fact that the reciprocity law[10] was used in writing Eq. (5.1). Accordingly, the conclusion that the photographic plate is a Type *II* detector is limited also by the condition that the durations of the exposure must be within the range over which the reciprocity law holds.

Some of the Eastman plates intended for use in spectroscopy and astronomy obey the reciprocity law fairly closely over an exposure range of 10^7 to 1.

Presumably situations exist in which the time constants relating to the formation and fading of the latent image are the controlling ones. It has not been shown that the conclusions of this section apply to these cases.

The writer wishes to express his gratitude to Dr. David L. MacAdam of the Eastman Kodak Company for constructive criticism of several earlier drafts of this section.

6. Photoelectric Detectors

This category is intended to include both gas and vacuum photo-electric tubes, and photomultiplier tubes.

The background which limits the sensitivity of these devices is the shot noise associated with the dark current of the photo-cathode. In order to obtain high sensitivity in the visible part of the spectrum, it is necessary to employ materials in the photo-cathodes which have a relatively low work function. Unfortunately, all such materials have a measurable thermionic emission at room temperature. Such emission may frequently be

[8] For a summary of the experimental information, See L. A. Jones and G. C. Higgins, J. Opt. Soc. Am. 36, 203–227 (1946).

[9] Private communication from Dr. David L. MacAdam dated January 4, 1949.

[10] *Photographic Plates for Scientific and Technical Use*, sixth edition (Eastman Kodak Company, Rochester, 1948), p. 11.

reduced to a negligible amount by cooling the tube to dry ice or liquid air temperature.[11]

Vacuum Photo-Tube

Consider first a simple vacuum photo-tube. Since the thermionic emission from the photo-cathode is temperature limited, the mean square fluctuation in the cathode current is given by the usual expression[12] for shot noise:

$$i^2 = 2eI_c\Delta f, \tag{6.1}$$

where I_c is the cathode dark current, e is the charge of the positron, and Δf is the noise equivalent band width.

This noise combines additively with the Johnson noise current of the load resistance R of the photo-tube:

$$i^2 = (4kT/R)\Delta f. \tag{6.2}$$

The total noise current is thus given by

$$i^2 = (2eI_c + 4kT/R)\Delta f. \tag{6.3}$$

It is evident from this relation that the noise of the load resistor may always be made small compared with the shot noise by making the load resistance R large enough. The condition is

$$I_cR \gg (2kT/e) = 0.0518 \text{ volt at } T = 300°\text{K}. \tag{6.4}$$

Thus in order to prevent the Johnson noise of the load resistance from being greater than the intrinsic noise of the photo-tube, it is necessary that the load resistance be chosen so that voltage drop produced in it by the cathode current is large compared with 0.0518 volt.

The magnitude of the required resistance may be estimated as follows. The cathode dark current of a very good photo-tube is roughly 10^{-15} ampere.[13] In order that this current produce a voltage drop equal to 0.05 volt, the load resistor must be 5×10^{7} megohms. If the anode capacity to ground has the value five micromicrofarads, the RC time constant is then 250 seconds. Since the anode capacity to ground cannot be less than the figure given, it follows that the physical time constant of a good photo-cell cannot be made less than about 250 seconds when it is operated in such a way as to obtain the minimum noise power per unit band width.

The physical time constant in this case is not equal to the reference time constant, since the noise power per unit band width is inversely proportional to the frequency squared for angular frequencies greater than 0.004 radian/sec. Indeed, the reference time constant is infinitely short, since the responsivity squared and the noise power per unit band width both decrease in the same way as the frequency is increased.

Accordingly, it is again necessary to introduce a "practical consideration." First of all, because of the extremely high impedance of the detector at low frequencies, the only practical way of amplifying the output of the photo-tube is to connect it to the grid of a vacuum tube amplifier. Furthermore, in practice, amplifiers themselves possess a certain amount of noise; this noise may be defined in terms of the input resistance whose Johnson noise would produce the same noise level at the output of the amplifier; this noise equivalent resistance R_{eq} is of the order of 10^5 ohms for conventional low frequency amplifiers. From this consideration it follows that the reference time constant is shorter than the RC time constant by a ratio equal to the square root of R_{eq}/R:

$$\begin{aligned} \tau &= R_0C(R_{eq}/R_0)^{\frac{1}{2}}, \\ \tau &= (R_0R_{eq})^{\frac{1}{2}}C, \end{aligned} \tag{6.5}$$

where R_0 is the load resistance which makes Eq. (6.4) an equality, R_{eq} is the noise equivalent resistance of the amplifier, and C is the sum of the capacity of the photo-tube and of the input of the amplifier.

Equation (6.5) indicates the shortest reference time constant attainable with a simple photo-tube if one requires that the shot noise of the photo-tube itself should be dominant. With the values $R_0 = 5\times10^{13}$ ohm, $R_{eq} = 10^5$ ohm, and $C = 5\times10^{-12}$ farad, this equation yields $\tau = 0.01$ sec.

Provided that the second term on the right-hand side of Eq. (6.3) is small compared with the first—that is to say, provided that the reference time constant is longer than that defined by Eq. (6.5), the noise equivalent photo-current is given by

$$\begin{aligned} i_n &= (2eI_c\Delta f)^{\frac{1}{2}} \\ &= (eI_c/2\tau)^{\frac{1}{2}}. \end{aligned} \tag{6.6}$$

Since the dark current I_c is proportional to the area of the photo-cathode, and since for a given photo-emissive surface the noise equivalent power is proportional to the noise equivalent current, it follows that the simple photo-tube is a Type I

[11] Ralph W. Engstrom, J. Opt. Soc. Am. **37**, 420–431 (1947).

[12] E. B. Moullin, *Spontaneous fluctuations of voltage* (Clarendon Press, Oxford, 1938), Chapter 2.

[13] Dr. Ralph Engstrom in a private communication dated January 3, 1949, suggests that the cathode dark current of the very best 1P21 is about 10^{-15} ampere, and that the average cathode dark current for this tube is about 10^{-14} ampere. The HB-3 RCA Tube Manual states that the maximum anode dark current of the 1P21 is 10^{-7} ampere when the gain is 2×10^6; this corresponds to a maximum cathode dark current of 5×10^{-14} ampere.

detector for reference time constants longer than that given by Eq. (6.5).

Gas Photo-Tubes and Photo-Multiplier Tubes

The discussion so far in this section has brought out the two fundamental difficulties in obtaining high signal-to-noise ratios with simple photo-tubes:

(1) The values of the minimum load resistance are fantastically high: about 10^8 megohms.
(2) The shortest time constant is about 0.01 second, which corresponds to a band width of 16 c.p.s.

The successful outcomes of the efforts to overcome these two difficulties were first, the gas photo-tube, and second, the photo-multiplier tube. The two are conveniently treated together.

Both contain internal amplification by means of which on the average G electrons reach the anode for each electron emitted by the photo-cathode. The amplification G is 10 for a typical gas photo-tube, and 10^6 for a typical photo-multiplier. The gas photo-tube suffers from the limitation that its amplification drops off above about 10^4 c.p.s., whereas the photo-multiplier maintains its gain up to about 10^9 c.p.s.,[14] at which point variations in electron transit time serve to reduce its response.

The mean square fluctuation in the anode current for these tubes is given by

$$\begin{aligned} i^2 &= 2eI_cFG^2\Delta f \\ &= 2eI_aFG\Delta f, \end{aligned} \tag{6.7}$$

where I_c is the cathode current, I_a is the anode current, and F is a numerical constant which takes account of the fact that the amplification process itself introduces noise which is over and above the amplified noise in the cathode current.

Let p_n be the probability that when one electron leaves the cathode, exactly n electrons reach the anode. Then F is defined by

$$F \equiv \frac{\sum_n n^2 p_n}{(\sum_n n p_n)^2}. \tag{6.8}$$

It is evident from the form of this expression that F is not less than unity. Various reasonable assumptions about the relative probabilities p_n lead to values of F which lie between unity and two.[15] The factor F will be omitted in the remainder of this paper.

[14] R. D. Sard, J. App. Phys. **17**, 768–777 (1946).

[15] J. A. Rajchman and R. L. Snyder, Electronics **13**, 20–23, 58, 60, December (1940); W. Shockley and J. R. Pierce, Proc. Inst. Rad. Eng. 26, 321–332 (1938); private communication dated August 17, 1948, from Dr. Hartland S. Snyder (Brookhaven National Laboratory).

The relation which now takes the place of Eq. (6.4) is

$$I_cR \gg 2kT/eG^2. \tag{6.9}$$

With the gas photo-tube and its gain G of 10, it follows from the numerical assumptions used earlier in this section that the minimum value of the load resistor is 5×10^5 megohms, the RC time constant is 2.5 seconds, and the shortest reference time constant is 0.001 second, or a band width of 160 c.p.s. On the other hand, with a photo-multiplier and its gain of 10^6 or more, the minimum value of the load resistor as given by Eq. (6.9) is only 50 ohms! With a capacity of 5×10^{-12} farad, the RC time constant is 2.5×10^{-10} second, which corresponds to a band width of 6×10^8 c.p.s. Thus with a photo-multiplier it is possible to obtain a band width of 600 megacycles, which band width is of the same order as that of the widest amplifier yet constructed.

In summary, all three detectors are Type I detectors for time constants greater than a certain minimum value. The minimum time constants are roughly 0.01 second, 0.001 second, and 2.5×10^{-10} second, for vacuum photo-tubes, gas photo-tubes, and photo-multiplier tubes, respectively.

The writer would like to express his appreciation of very helpful criticism of an early draft of this section, rendered independently by: Dr. Ralph Engstrom, Dr. J. A. Rajchman, and Dr. Albert Rose, all of Radio Corporation of America, and Dr. Hartland S. Snyder, Brookhaven National Laboratory.

7. The Dipole Antenna

The dipole antenna is a somewhat different type of detector from the others which are considered in this paper, and the treatment of it is somewhat artificial from the present point of view.

There are three ways in which the performance of the dipole antenna differs from those treated so far in this paper:

1. The output voltage is not proportional to the power incident upon its effective area, but it is rather the output power which is proportional to the incident power.

2. The directivity pattern is different from that of the other detectors. More importantly, if one varies the area of the antenna by the only available method, namely by varying the number of dipoles in an array, the noise equivalent power is not proportional to the square root of the effective area, but is rather independent of the effective area. This fact is tied in with the fact that the directivity of the detector increases as the effective area increases. This difficulty may be avoided simply by introducing instead of the area, the product of the area and the reciprocal directivity index. This possibility has not been developed in these papers, however, because our primary interest lies in detectors which obey Lambert's law.

3. Temperature noise and Johnson noise become identical. From one point of view, the noise at the output of a dipole

antenna may be considered as the Johnson noise associated with the radiation resistance of the dipole, but from another point of view the noise may be considered as the response of the antenna to the thermal radiation field in which it is placed. The same result is obtained from either point of view. This situation is due to the fact that, except for ohmic losses, the dipole antenna is a completely reversible transducer with an efficiency of unity.

If, however, one wishes to force the dipole antenna into the mold established in Part I, it may be done as follows.

In an array of dipoles, let each dipole be connected separately through a matched transmission line to a square law rectifier, and let the outputs of the rectifiers be connected in series. It is supposed that the input of the rectifier has an impedance high compared with that of the dipole. Let ν denote the frequency of the voltage at the input of the rectifier, and let f denote the frequency at the output. Then the behavior of the square law rectifier may be written

$$V_f = \alpha V_\nu^2, \tag{7.1}$$

where V is the instantaneous voltage.

For the sake of simplicity of treatment, it is further supposed that the highest frequency f in the output of the rectifier is small compared with the lowest frequency ν at the input. It is further assumed that the band width Δf at the output of the rectifier is small compared with the band width $\Delta\nu$ of the dipole. Then it is easily shown that the noise power per unit band width N_f in the output of the detector is related to that at the input by

$$N_f = \alpha^2 \Delta\nu N_\nu^2, \tag{7.2}$$

where N_ν is assumed to be constant within the band width $\Delta\nu$, and zero otherwise. The noise power per unit band width N_ν is simply the Johnson noise associated with the radiation resistance Z of the dipole

$$N_\nu = 4kTZ, \tag{7.3}$$

whence

$$N_f^{\frac{1}{2}} = 4\alpha kTZ(\Delta\nu)^{\frac{1}{2}}. \tag{7.4}$$

Let the response to a signal be considered next. If P is the signal power incident upon the effective receiving area of the antenna, the open-circuit signal voltage at the input of the rectifier is given by

$$V_\nu^2 = ZP, \tag{7.5}$$

where Z is the radiation resistance of the dipole. One then has by Eq. (7.1)

$$V_f = \alpha ZP \tag{7.6}$$

or

$$S \equiv V_f/P = \alpha Z. \tag{7.7}$$

By combining Eqs. (7.4) and (7.7) one finds that the zero-frequency responsivity-to-noise ratio for a single dipole is

$$\mathfrak{R}_0 = \frac{S}{N_f^{\frac{1}{2}}} = \frac{1}{4kT(\Delta\nu)^{\frac{1}{2}}}. \tag{7.8}$$

If the dipoles in the array are separated by a distance of at least one-half wave-length, so that the noises at their outputs are nearly completely independent in the statistical sense, then one may use the argument given in Part I, Section 5, with the result that $\mathfrak{R}_0$ for the array is inversely proportional to the square root of the number of dipoles, and thus also inversely proportional to the square root of the total area of the array.

Thus, since $\mathfrak{R}_0$ is independent of the reference time constant, and is inversely proportional to the square root of the area, it follows that the array with the associated square law detectors is a Type *I* detector.

The argument given above is interesting in that no attempt was made to define a reference time constant. It was merely shown (Eq. (7.8)) that for frequencies f within the output pass-band of the rectifier, the responsivity-to-noise ratio is independent of the width of this band.

8. Other Detectors

The most important detector which has been omitted from discussion is the human eye. The eye, however, cannot be discussed fruitfully from the present point of view, because neither the area nor the time constant of the eye are adjustable over any substantial range. The eye is most usefully discussed on the basis of the third type of noise background defined in Section 2 of Part I. This treatment has been carried out by Albert Rose.[3]

Perhaps the next most important detector which has been omitted is the photo-conductive cell. In this case, the detector has not yet been developed from either the theoretical or the experimental aspect, to the point where one knows the dependence of the noise equivalent power on the reference time constant. The writer considers it highly probable that the photo-conductive cell is either a Type *I* or a Type *II* detector. Much the same statements may be made about the photovoltaic cell.

No effort has been made in this paper to consider the detectors for x-rays, γ-rays and high energy

particles such as ionization chambers, proportional counters, Geiger counters, et cetera.

ACKNOWLEDGMENT

The writer wishes to express his appreciation to Dr. Marcel J. E. Golay of the Signal Corps, U. S. Army, for a helpful discussion of this method of classification.

Erratum: The Ultimate Sensitivity of Radiation Detectors

[J. Opt. Soc. Am. 37, 879 (1947)]

R. CLARK JONES

Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts

THE writer has had the privilege of seeing in advance of publication a manuscript by Mr. P. B. Fellgett, Peterhouse, Cambridge, England. In the manuscript, Mr. Fellgett points out that the expression (3.8) in this paper holds less generally than is claimed in the paper. Although the expression holds for the case for which it was derived—namely, for detectors whose response is mediated by a change in temperature of the sensitive element, it is incorrect in application to a photo-cell for two reasons:

1. In the heat detector, the fluctuations are due equally to the radiation incident upon, and emitted by, the detector. But in a photo-cell, only the fluctuations in the incident radiation are effective. This change has the effect of multiplying the noise equivalent power by $2^{-\frac{1}{2}}$.

2. In the heat detector, the photons are weighted in accordance with their energy, whereas in the case of the photo-cell, the photons are weighted in accordance with their number. This change requires that for a given quantum efficiency, the weighting of photons of different frequency must be done differently. Evidently, however, if the fractional band width of the detector is small, this change is unimportant.

Mr. Fellgett concludes that his observations serve to modify the expression (3.8) by numerical factors whose order of magnitude is unity in all practical cases.

The writer wishes to thank Mr. Fellgett for communicating a copy of his manuscript in advance of publication and for approving the publication of these remarks.

The author wishes to take this opportunity to correct several misprints of minor importance in Paper I, all on page 886:

In the legend of Fig. 1, the temperature *300*°F should be *300*°K, and the word *derector* should be *detector*.

In column two, the designation *300*K should be *300*°K.

The last equation on page 886 should bear the number 9.2.

Factors of Merit for Radiation Detectors*

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts

(Received February 4, 1949)

In Part I, the need for a suitably defined factor of merit for comparing the sensitivity of different radiation detectors is explained (Section 1). Both the advantages and the very important limitations of such a factor of merit are discussed.

In Section 2, separate factors of merit for Type *I* and Type *II* detectors are defined. The factor of merit for Type *I* detectors is defined as the ratio of the noise equivalent power given by the fundamental limit defined in Paper I to the noise equivalent power actually measured in the reference condition. The factor of merit for Type *II* detectors is similarly defined in terms of the actual noise equivalent power in the reference condition and the empirically determined minimum value for the noise equivalent power proposed by R. J. Havens. Section 3 contains a brief discussion of the significance of the factor of merit for the special case of the bolometer.

In Part II, the noise equivalent power in the reference condition, and the factor of merit is determined for about fifty thermocouples and bolometers. The results are presented in Tables I and II, and in Figs. 1 and 2. A description of each detector is given in Section 5, along with a statement of any special calculations which were necessary.

All of the results presented are based on measurements made by other persons; no measurements were performed by the writer. The sources of information are described in Section 3.

An Appendix is provided whose purpose is to call attention to the wide variety of possibilities in the specification of the relevant properties of radiation detectors, and thereby to encourage precision in the statement of these properties in the future periodical literature.

PART I. DEFINITION OF THE FACTORS OF MERIT

1. Introduction

THERE exists a need for a single, quantitative factor of merit for use in comparing the sensitivity of various radiation detectors. This criterion is needed for the intercomparison of similar detectors, such as an evaporated thermocouple and a wire thermocouple, and also for the intercomparison of dissimilar detectors, for example, a Golay pneumatic heat detector and a lead sulfide photoconductive cell.

Such a factor of merit should be capable of intercomparing units with sensitive areas of greatly different magnitude, with different speeds of response, and with different spectral response curves, when measured with an amplifier having any given frequency-response curve.

No such criterion has been developed in the past. In a few cases where the intercomparison of a number of detectors has been attempted, the "comparison" has consisted in the presentation of a table with a large number of entries for each detector, no one of which entries by itself permits a direct comparison of one detector with another.

Presumably the reasons for the failure to compare detectors in a way which, in however restricted a sense, permits a statement of a numerical factor of relative merit, are that

(1) comparisons are unpopular,

(2) comparisons carry with them a heavy load of responsibility, and

(3) a single factor of merit cannot take into account a number of important secondary considerations, such as ease of amplification, over-all physical size, need for auxiliary facilities, the physically attainable ranges of sensitive area and speed of response, and also price and delivery time.

Nevertheless, when it is necessary to choose a detector for a given task, a decision must be made. The writer proposes to assist in this decision by defining a factor of merit which depends upon, and only upon, the relative signal-to-noise ratios obtainable with the various detectors.

This paper contains the writer's proposal of a numerical factor of merit. The proposed factor of merit takes no account of the secondary considerations mentioned above as item 3, and the writer accepts the implications of items 1 and 2. It is offered not as a final judgment, but as a beginning, to be tempered and modified in the light of experience and opinion.

2. Definition of the Factor of Merit

The factor of merit proposed here is based on the reference condition of measurement defined in Paper II and, indeed, is almost implicit in the definition of the classification system. In Part I, Section 9 of Paper II a Type n detector is defined as one whose noise equivalent power P_m in the reference condition A depends upon the reference time constant τ and the sensitive area A according to

$$P_m=(A^{\frac{1}{2}}/k_n\tau^{\frac{1}{2}n}), \tag{2.1}$$

where k_n is a parameter which is independent of A and τ, but which has different values for different detectors.

Since for a given area and time constant it is desired to make the noise equivalent power as small

* This is the third of a series of papers, the second and third of which appear in this issue. The work reported in this paper was supported in part by U. S. Navy Contract NObsr-42179.

as possible, it is evident that k_n is itself a factor of merit for a Type n detector. This parameter is of inconvenient magnitude for ordinary purposes, however, and one accordingly looks for a suitable way of defining a numerical multiple of k_n as the factor of merit.

For Type I detectors, a convenient multiple may be obtained by using Eq. (3.8) of Paper I. For a detector whose quantum efficiency is unity at every radiation wave-length, Eq. (3.8) states that the minimum value of the noise equivalent power is given by

$$P_m=(4\sigma T^4kT)^{\frac{1}{2}}(A^{\frac{1}{2}}/\tau^{\frac{1}{2}}), \tag{2.2}$$

where k is Boltzmann's constant, σ is the Stefan-Boltzmann radiation constant, and where T is the absolute temperature of the detector and of the surrounding radiation field. At the temperature $T=300°\text{K}$ the last equation may be written

$$P_m=2.76\times10^{-12}(A^{\frac{1}{2}}/\tau^{\frac{1}{2}}), \tag{2.3}$$

where P_m is in watts, A is in square millimeters, and τ is in seconds. If a detector satisfying Eq. (2.3) be considered to have a factor of merit equal to unity, then the factor of merit for any other Type I detector may be written

$$\begin{aligned} M_1&=2.76\times10^{-12}k_1,\\ &=2.76\times10^{-12}(A^{\frac{1}{2}}/P_m\tau^{\frac{1}{2}}),\\ &=5.52\times10^{-12}\mathfrak{R}_0A^{\frac{1}{2}}, \end{aligned} \tag{2.4}$$

where P_m is in watts, A is in square millimeters, τ is in seconds, k_1 is in the units resulting from using the units just mentioned in Eq. (2.1), and $\mathfrak{R}_0$ is in watt^{-1}-sec.$^{-\frac{1}{2}}$.

Equation (2.4) is the proposed factor of merit for a Type I detector.

For Type II detectors, a convenient multiple may be obtained by using Havens' limit. About three years ago, R. J. Havens** made an estimate of the minimum value of the noise equivalent power which could be obtained with thermocouples and bolometers with currently available materials and techniques, when the detector was at room temperature. His estimate was in no way intended to be a fundamental limit; it has the nature rather of an optimistic engineering estimate. His estimate has been very well confirmed in the sense that no bolometer or thermocouple known to the writer which operates at room temperature has better performance than is indicated by his estimate, although several detectors have approached it within a factor of two, as is shown in Part II.

Havens' limit is

$$P_m=3\times10^{-12}(A^{\frac{1}{2}}/\tau), \tag{2.5}$$

where P_m is in watts, A in square millimeters, and τ in seconds. If a detector which satisfies Eq. (2.5) be considered to have a factor of merit equal to unity, then the factor of merit M_2 for any other Type II detector may be written

$$\begin{aligned} M_2&=3.0\times10^{-12}k_2,\\ &=3.0\times10^{-12}(A^{\frac{1}{2}}/P_m\tau),\\ &=6.0\times10^{-12}(\mathfrak{R}_0A^{\frac{1}{2}}/\tau^{\frac{1}{2}}), \end{aligned} \tag{2.6}$$

where P_m is in watts, A is in square millimeters, τ is in seconds, k_2 is in the units resulting from using the units just mentioned in Eq. (2.1), and $\mathfrak{R}_0$ is in watt^{-1}-sec.$^{-\frac{1}{2}}$.

Equation (2.6) is the proposed factor of merit for a Type II detector.

The factor of merit M_2 for about fifty bolometers and thermocouples is plotted in Fig. 2.

3. Significance of the Factor of Merit

Because of the rather formal definition of the factors of merit given above, it seems appropriate at this point to discuss briefly the significance of the factor of merit.

In order to make the discussion as specific as possible, it will be supposed that the detector in question is a bolometer limited by Johnson noise. As shown in Part II, Section 2 of Paper II, the thermal capacity per unit area is to be held at its minimum attainable value (which value is supposed to be independent of the sensitive area) as the sensitive area is changed. The reference time constant is changed by varying the effective cooling constant κ, while again the thermal capacity per unit area is held constant at its minimum value. On the basis of this prescription, the factor of merit M_2 of the bolometer remains constant as the sensitive area and reference time constant are changed.

Consider now two bolometers with different sensitive areas and different time constants but with the same factor of merit. Because of the difference in their sensitive areas and time constants these two bolometers, in spite of the fact that their factors of merit are the same, will not in general yield equivalent results in any *particular* application. The fact that the two bolometers have the same factor of merit means, however, that if the two bolometers are reconstructed so that they each have the optimum sensitive area and optimum time constant for the particular application, then the performance of the two will be the same.

If the two bolometers have different factors of merit, then when they are each constructed so that they have the optimum sensitive area and time constant for a given application, the signal-to-noise ratios will be in proportion to their factors of merit.

** R. J. Havens, J. Opt. Soc. Am. 36, 355A (1946).

PART II. FACTORS OF MERIT FOR FIFTY RADIATION DETECTORS

1. Introduction

This part contains the calculation of the factors of merit for about fifty Type *II* radiation detectors. The results of the calculations are shown in Tables I and II, and in Figs. 1 and 2.

As described in Section 3, a large number of references was consulted in connection with the collection of the data presented here. Unfortunately, a substantial number of the detectors described in the periodical literature are not described adequately for the present purpose, with the result that it was not possible to extract the information needed to calculate the factor of merit.

In some cases the missing information was obtained by correspondence with the authors, and in other cases the detector was omitted from consideration in this paper.

The most common omission in the published descriptions is information about the noise equivalent band width employed in making a measurement of the signal-to-noise ratio. In the older literature, information about the detector is often inextricably mixed with the properties of the associated galvanometer.

Because of the frequency of incomplete or vague description in the periodical literature, the writer has included in an Appendix to this paper a number of suggestions with regard to the presentation of information about measurements on detectors.

2. Assumed Operating Conditions

In computing the noise equivalent power in either of the reference conditions defined in the preceding paper, it is necessary to adopt several arbitrary conventions.

In the case of all detectors, it is assumed that they work into an open circuit, rather than into a load resistance equal in value to the resistance of the detector. In the latter case the factor of merit is less by the factor $2^{-\frac{1}{2}}$ than in the former case, if the only noise is the Johnson noise of the detector.

In the case of bolometers, it is assumed that the bolometer is operated with a series resistance large compared with the resistance of the bolometer strip, so that the responsivity of the detector in volts per watt is the product of the resistance change in ohms per watt multiplied by the maximum permissible current through the bolometer.

In the case of thermocouples and thermopiles, it is assumed that the construction of the hot and cold junctions is symmetrical, so that the resistances associated with the two types of junctions are equal. The Johnson noise is calculated for the series resistance of the hot and cold junctions.

In one sense, these conventions discriminate against the thermocouples in favor of the bolometers, since in principle one may use a cold junction of very low resistance, and thus obtain a gain in signal to Johnson noise ratio by the factor $2^{\frac{1}{2}}$ over that involved in the above convention with regard to thermocouples. On the other hand, however, in practice one nearly always uses a cold junction equal in resistance to that of the hot junction, whereas with bolometers one often uses a series resistance several times the bolometer resistance. Anyone who disagrees with the propriety of these conventions may divide the factor of merit of the bolometers by the factor $2^{\frac{1}{2}}$ on the basis that they should be used with a series resistance equal to the bolometer resistance, or he may increase the factor of merit of the thermocouples by the factor $2^{\frac{1}{2}}$ on the basis that they may be used with low resistance cold junctions.

3. Sources of Information

The data used in this report were obtained from three sources: the periodical literature, private communications, and Professor Harald H. Nielsen's comprehensive OSRD report[1] on the Ohio State tests of thermal detectors.

The available information was divided into two classes. Table I includes those detectors for which actual noise measurements were not available, and for which accordingly it was necessary to assume that the noise could be calculated on the basis that it was all Johnson noise. For the detectors in Table I the noise equivalent power and the factor of merit were calculated by Eqs. (4.5) and (4.7) below.

Table II includes those detectors for which actual noise measurements were available, so that it is possible to calculate the noise equivalent signal in the presence of the actual noise, rather than the noise calculated by assuming that the only noise is Johnson noise. For the detectors in Table II, the noise equivalent power and the factor of merit were calculated by Eqs. (4.9) and (4.10).

Although the actual noise in the output of thermocouple detectors is usually no higher than the computed Johnson noise, the actual noise of bolometers is frequently higher, sometimes much higher, than the computed Johnson noise. Accordingly, it should be recognized that the actual factors of merit of detectors 23 through 33 are probably somewhat less than those presented in Table I and plotted in Fig. 2. The degree to which they are less is not determinate on the basis of information now available.

There is a similar weakness in the case of some of the entries in Table II. In some cases, the noise was

[1] Harald H. Nielsen, *Comparative Testing of Thermal Detectors*, OSRD Report No. 5992, Contract OEMsr-1168 (October 31, 1945).

measured in a bandwidth which is much wider than that of the detector. This has the result that if the noise is high within the band width of the detector, but is otherwise lower, the high noise within the detector band width does not appear in the result of the measurement.***

Professor Nielsen's group actually measured the noise of the detectors studied by them. Unfortunately, however, as is emphasized in their report, the amplifiers used were not characterized by an unusually low noise level, with the result that the measured noise often reflected the noise of the amplifier rather than the noise of detectors. Since the relative importance of the noise added by the amplifiers varied from unit to unit, depending on the impedance of the detector, it was felt that the use of the measured noise would not permit an altogether equitable intercomparison of the different units. Accordingly, none of the noise measurements reported by Nielsen have been used, and all of the units tested by Nielsen are therefore included in Table I.

Although in a few cases (Numbers 22 and 46, for example) the data represent the typical performance of the manufacturer's production, most of the entries in Tables I and II represent measurements made on a single detector. Furthermore, in some cases (Numbers 3 and 23, for example), only a single sample of the detector has been tested. It is evident that conclusions based on such information should be accepted with some reserve.

Much of the information in the tables is not up-to-date. In particular, all of the tests reported by Nielsen[1] were completed before October 31, 1945. It is reasonable to suppose that in the meantime many of the detectors have been improved.

No information is given here about the spectrum of the radiation used to measure the noise equivalent power. This information is lacking in many of the references used, and is useful only if one knows also the response of the detector as a function of wave-length.

4. Derivation of Required Formulas

For those detectors listed in Table I, it is assumed that the only source of noise is the Johnson noise associated with the resistance of the detector. In this case simple expressions for the noise equivalent power and the factor of merit may be given in terms of the area A, the resistance R, the effective time constant τ, and the effective zero-frequency responsivity S_0.

*** The writer wishes to acknowledge his indebtedness to Dr. J. A. Becker, Bell Telephone Laboratories, for pointing out the need for emphasizing the weakness in the data mentioned in this paragraph and the preceding one, and also for a helpful discussion of the sources of noise in bolometers.

The signal voltage E_s obtained from a given steady power P incident upon the detector is given by

$$E_s = S_0 P. \tag{4.1}$$

The r.m.s. Johnson noise voltage delivered by the detector is given by

$$E_N = (4kTR\Delta f)^{\frac{1}{2}} = (kTR/\tau)^{\frac{1}{2}}, \tag{4.2}$$

where k is the Boltzmann gas constant. The noise equivalent power P_m in the reference condition A is now to be obtained by equating these expressions. One finds

$$P_m = (kTR/\tau)^{\frac{1}{2}}/S_0. \tag{4.3}$$

Upon reducing the noise equivalent power to a sensitive area of 1 mm² by the use of Eq. (6.6) of Paper II, Part I, one finds that the noise equivalent power H_m in the reference condition B is given by

$$H_m = (kTR/A\tau)^{\frac{1}{2}}/S_0, \tag{4.4}$$

where A is the sensitive area in square millimeters. If the detector is at a temperature of 300°K, Eq. (4.4) becomes

$$H_m = 0.64\times10^{-10}(R/A\tau)^{\frac{1}{2}}/S_0, \tag{4.5}$$

where H_m is in watts, R is in ohms, A is in square millimeters, τ is in seconds, and S_0 is in volts/watt.

The factor of merit M_2 for Type *II* detectors now takes the form

$$M_2 = 3\times10^{-12}S_0(A/kTR\tau)^{\frac{1}{2}}, \tag{4.6}$$

or, when the detector is at a temperature of 300°K,

$$M_2 = 0.0468S_0(A/R\tau)^{\frac{1}{2}}, \tag{4.7}$$

where the units are the same as in Eq. (4.5). Thus, in order to determine the factor of merit for a detector whose limiting sensitivity is set by Johnson noise alone, the factor of merit may be defined in terms of the effective zero-frequency responsivity S_0, the sensitive area A, the resistance R, and the reference time constant τ.

For those detectors which are listed in Table II, the noise was actually measured. The entries tabulated in Table II for each detector are defined as follows.

The power P_0 denotes the value of the steady incident power which produces a steady output voltage equal to the noise voltage under the actual condition of measurement. If the measurement was actually made with a sinusoidally modulated signal, the result tabulated was obtained by reducing the measured result to zero frequency by making use of the measured frequency response curve of the de-

tector. Furthermore, if the detector does not possess a single time constant, the power P_0 is the effective value, as defined in Section 7, Part I, of Paper II.

The quantity Δf is the noise equivalent band width in the actual measurement, and A and τ denote the sensitive area and the reference time constant.

Then, since the noise equivalent band width in the reference condition is $1/(4\tau)$, the noise equivalent power P_m in the reference condition A is related to the noise equivalent power P_0 by

$$P_m = P_0/(4\tau\Delta f)^{\frac{1}{2}}. \tag{4.8}$$

The noise equivalent power H_m in the reference condition B is

$$H_m = P_0/(4\tau A\Delta f)^{\frac{1}{2}}. \tag{4.9}$$

From the last expression the factor of merit M_2 for Type *II* detectors is given by

$$M_2 = 6\times 10^{-12}(A\Delta f/\tau)^{\frac{1}{2}}/P_0, \tag{4.10}$$

where P_0 is in watts, A is in square millimeters, Δf is in cycles/second, and τ is in seconds.

5. Description of the Detectors

This section contains a brief description of each detector, and a statement of the source of information. If any intermediate computation was necessary, the nature of the calculation is stated.

The numbers in parentheses after the name of each detector give the reference numbers used in the tables and figures.

The letter N in parentheses indicates that the data were obtained from Nielsen's report.

The Weyrich Vacuum Thermocouple (1–2) (N)

This thermocouple has been used for many years by infra-red spectroscopists. It may be used at atmospheric pressure or it may be evacuated. The thermoelectric materials are two different antimony-bismuth alloys. Both of the junctions of the thermocouple are designed for use as radiation receivers. The values entered in Table I refer to the area of a single receiver and the resistance of the two junctions in series. The time constants were obtained with the use of a ballistic galvanometer.

The Eppley (Farrand) Thermocouple (3) (N)

This thermocouple was developed by the Eppley Laboratory and is of the type supplied for use in the Farrand heat detector. As with the Weyrich thermocouple, both of the junctions are fitted with radiation receivers. The information in Table I refers to the area of a single receiver and to the resistance of the two junctions in series. The time constant was determined from the frequency response curve, since it is of the form which corresponds to a single time constant.

The Schwarz Thermopiles (4–6) (N)

These thermopiles are manufactured by Adam Hilger, Ltd. of London, and employ a pin type of construction. The thermopiles contain two hot receivers and two cold receivers, each 2 mm² in area. The information in Table I refers to the area of two receivers and to the resistance of the four receivers in series. The measured zero-frequency responsivities of the lower pair of receivers for the three units listed in Table I are 0.75, 1.00, and 1.2 volt/watt.

Information with regard to the frequency response of these thermopiles is given only for unit No. 11. The frequency response curve of these units is of Class *I*, defined in Section 7, Part I of Paper II. The curve has a slope of 3 db per octave at 8 c.p.s., at which point the responsivity is 8.2 db less than it is at zero frequency. Accordingly, the reference time constant is 20 milliseconds and the effective zero-frequency responsivity is 5.2 db less than the measured zero-frequency responsivity.

If one assumes that the relative frequency response curve of units No. 9 and No. 12 is the same as that of No. 11, one finds that the effective zero-frequency responsivities are 0.41, 0.55, and 0.66 volt/watt, respectively.

The Buck-Harris Evaporated Thermopiles (7–8) (N)

These thermopiles were developed by Willard E. Buck and Dr. Louis Harris at the Massachusetts Institute of Technology. The units were prepared by an evaporation technique and were of the folded type. The case is not evacuated. The values entered in Table I refer to the area of the hot receivers and to the resistance of the entire thermopile.

The measured zero-frequency responsivity of the two units listed in Table I are 0.345 and 0.375 volt/watt.

The frequency response curve of these detectors is of Class *I*, defined in Part I of Paper II. Unit No. 1 has a slope of 3 db per octave at 12 c.p.s., at which point the response is 4.8 db less than its zero-frequency value; the reference time constant and effective zero-frequency responsivity are thus 13.3 milliseconds and 0.280 volt/watt. The frequency response curve of unit No. 2 has a slope of 3 db per octave at a frequency of 12 c.p.s., at which point the response is 4.0 db less than at zero-frequency. Accordingly, the reference time constant is 13.3 milliseconds and the effective value of the zero-frequency responsivity is 0.334 volt/watt.

The Eppley (Emerson) Thermopile (9) (N)

This thermopile was manufactured by Eppley Laboratory for the Emerson Radio and Phonograph Corporation. The thermopile contains 12 hot junctions and 12 cold junctions arranged so that the 12 hot receivers cover an area $\frac{1}{4}'' \times \frac{1}{2}''$, and so that the cold receivers fill a similar adjacent area, with the result that the receivers cover an area $\frac{1}{2}'' \times \frac{1}{2}''$. The information in Table I refers to the area of the hot junctions and to the resistance of all of the junctions. The time constant was determined by using a ballistic galvanometer.

The Harris Thermocouples (10–11)

These two thermocouples are described by Louis Harris, Phys. Rev. **45**, 635–640 (1934). Each is evacuated, and is backed with a cellulose film. In both cases, the effective area of the hot junction is 1.2 mm². The bismuth-tellurium couple has a resistance of 150,000 ohms, a time constant of 212 milliseconds, and a zero-frequency responsivity of 7.6×10^{-5} volt/(44 calories/(cm²/sec.)) = 34.2 volts watt. The bismuth-antimony couple has a resistance of 25,000 ohms, a time constant of 90 milliseconds, and a zero-frequency responsivity of 8.8×10^{-6} volt/(44 calories/(cm²/sec.)) = 3.96 volts/watt. The time constants are calculated from the fact that the attenuation factors at 30 c.p.s. are 0.025 and 0.059, respectively.

The Harris and Scholp Thermocouples (12–16)

These evacuated thermocouples are described by Louis Harris and Alvin C. Scholp, J. Opt. Soc. Am. **30**, 519–522 (1940); **31**, 25 (1941), supplemented by a private communication from Professor Harris dated May 28, 1948, and were made by the sputtering process. The units specified in Table I as items 12 through 16 are the blackened units designated in Table II of the reference cited above as couples 3, 5, 7, 8, and 2, respectively. All but No. 7 have an effective area of 1.2 mm², whereas No. 7 has an effective area of 4.8 mm². The first three couples employ bismuth and antimony, whereas the last two employ bismuth and tellurium. The authors state the zero-frequency responsivity in units of 10^{-5} volt, with an irradiation of 10^{-4} watt/cm²; their values have been converted to volts per watt by making use of the stated areas.

The Hornig and O'Keefe Thermopiles (17–19)

These are wire thermocouples employing a bismuth-antimony alloy as one conductor and a bismuth tin alloy as the other. They are described by D. F. Hornig and B. J. O'Keefe, Rev. Sci. Inst. **18**, 474–482 (1947). The information in Table I is taken directly from the authors' Table III.

The Perkin-Elmer Thermocouples (20–22)

These are evacuated thermocouples employing unspecified alloys assembled in a pin type of construction. The thermocouples were developed in the laboratories of the Perkin-Elmer Corporation. The thermocouples are described by Max D. Liston, J. Opt. Soc. Am. **37**, 515(A) (1947), and in the New Instruments Section, Rev. Sci. Inst. **18**, 373 (1947). The area of the receiver attached to the hot junction is 0.4 mm².

The unit described in the first reference has a resistance of 10 ohms, a zero-frequency responsivity of 5 volts/watt, and a speed such that the peak response to a square wave signal of 5 c.p.s. is 70 percent of the response to a steady signal.

In the second reference the resistance is specified as 20 ohms, the zero-frequency responsivity as 10 volts/watt, and a speed such that the peak response to a square wave signal of 5 c.p.s. is 80 percent of the response to a steady signal.

In a private communication to the writer dated June 1, 1948, Mr. Liston stated that the current production thermocouples had the same resistance and sensitivity as specified in the second reference but with an increased speed such that the peak response to a square wave signal of 15 c.p.s. is 75 percent of the response to a steady signal.

By use of the formula

$$\text{Attenuation Factor} = \tanh 1/(4\Delta f)$$

appropriate for peak response to square wave irradiation, one finds that the time constants to be associated with the three units are 58 milliseconds, 45 milliseconds, and 17 milliseconds, respectively.

The Strong Bolometer (23) (N)

This bolometer was constructed by Dr. John Strong in the laboratories of Harvard University. It contains six unbacked nickel strips operating in a hydrogen atmosphere at a pressure of 1 mm of mercury. The sensitivity is stated to be 2.54 ohms per incident watt, with a current of 0.24 ampere. This corresponds to a measured zero-frequency responsivity of 0.61 volt/watt.****

The frequency response curve of this detector is of Class *II*, defined in Paper II. The response at 30 c.p.s. is 3 db less than at zero frequency. The reference time constant is accordingly 5.3 milliseconds.

**** This is incorrectly given as 0.3 volt/watt in reference 1. (Confirmed by a private communication from Dr. E. T. Bell dated June 25, 1948.)

TABLE I. Data for the 33 detectors assumed to have a noise output equal to the calculated Johnson noise.

Reference number	Alternate identification	Time constant τ milliseconds	Zero-frequency responsivity S_0 volts/watt	Sensitive area A mm²	Resistance R ohms	Noise equivalent power H_m 10^{-10} watts	Factor of merit $1/M_2$	M_2
The Weyrich Vacuum Thermocouple (N)								
1	in air	68	0.29	2.0	20	26.7	60.5	0.0165
2	evacuated	1000	4.35	2.0	20	0.465	15.5	0.0645
The Eppley (Farrand) Thermocouple (N)								
3		90	0.375	1.0	5.8	13.7	41.1	0.0241
The Schwarz Thermopiles (N)								
4	#B4772/9	20	0.41	4.0	23.3	26.7	17.8	0.0562
5	#B4772/11	20	0.55	4.0	51.0	29.3	19.5	0.0513
6	#B4772/12	20	0.66	4.0	20.7	15.6	10.4	0.0962
The Harris Evaporated Thermopiles (N)								
7	#1	13.3	0.280	11.0	100	59.7	26.5	0.0377
8	#2	13.3	0.334	11.0	100	50.0	22.2	0.0450
The Eppley (Emerson) Thermopile (N)								
9		250	0.1	80	18	5.76	48.0	0.0208
The Harris Thermocouples								
10	Bi-Te	212	34.2	1.2	150,000	14.4	98.9	0.0101
11	Bi-Sb	90	3.96	1.2	25,000	77.7	233	0.00429
The Harris and Scholp Thermocouples								
12	#3	177	50	1.2	1480	1.07	6.31	0.166
13	#5	110	47.5	1.2	1430	1.40	5.13	0.195
14	#7	134	60.2	4.8	4900	0.928	4.15	0.241
15	#8	345	406	1.2	183,000	1.05	12.1	0.0826
16	#2	175	400	1.2	90,000	1.05	6.13	0.163
The Hornig and O'Keefe Thermopiles								
17		36	6.5	0.5	5	1.64	1.97	0.508
18		36	6.5	1.0	10	1.64	1.97	0.508
19		41	3.8	4.0	12.5	1.47	2.01	0.498
The Perkin-Elmer Thermocouples								
20	J. Opt. Soc. Am.	58	5	0.4	10	2.66	5.14	0.195
21	Rev. Sci. Inst.	45	10	0.4	20	2.13	3.20	0.313
22	letter	17	10	0.4	20	3.47	1.97	0.508
The Strong Bolometer (N)								
23		5.3	0.61	5.7	4.2	12.4	2.19	0.457
The Felix Bolometers (N)								
24	2A	20.2	1.09	17.2	16	3.99	2.69	0.371
25	15	21.2	1.41	17.2	16	3.01	2.13	0.469
26	11A	21.5	1.14	17.2	16	3.69	2.64	0.379
27	19	21.2	2.26	17.2	16	1.88	1.33	0.752
The Polaroid Bolometers (N)								
28	Ni 324	4.7	1.46	4.5	64	24.1	3.78	0.265
29	Ni 347	7.2	1.60	4.5	44	14.7	3.53	0.283
30	Ni 350	3.6	1.61	4.5	128	35.3	4.24	0.236
The Thermistor Bolometers (N)								
31	S-19	5.9	730	0.60	3×10^6	25.5	5.02	0.199
32	S-20	1.27	378	0.62	3×10^6	105.0	4.45	0.225
33	XB-108	135	3460	0.58	3×10^6	1.14	5.13	0.195

The Felix Bolometers (24–27) (N)

The Felix bolometers were developed by the Heat Research Laboratory of Massachusetts Institute of Technology under the direction of Alan C. Bemis. The bolometer contains four unbacked nickel strips operating in an atmosphere of hydrogen at a pressure of 4 mm of mercury. The measured zero-frequency sensitivities of the four bolometers tested, in the order used in Table I, are 10.3, 13.3, 10.7, and 21.2 ohms/watt. Since the current was 0.106 ampere, the measured zero-frequency responsivities are 1.09, 1.41, 1.14, and 2.26 volts/watt. The time constants were determined from the frequency response curves, since they are of the type which corresponds to a single time constant.

The Polaroid Bolometers (28–30) (N)

The Polaroid bolometers were developed in the laboratories of the Polaroid Corporation under the direction of Dr. Bruce Billings. These bolometers

TABLE II. Data for 15 detectors whose noise level was directly measured.

Reference number	Alternate identification	Time constant τ milliseconds	Noise equivalent power P_0 10^{-10} watt	Sensitive area A mm²	Band width Δf c.p.s.	Noise equivalent power H_m 10^{-10} watt	Factor of merit $1/M_3$	M_3
The Superconducting Bolometers								
36	22–*k*–½	10.0	3.67	1.8	4000	0.216	0.0720	13.9
37	25–*k*–½	1.2	14.5	1.2	4000	3.02	0.121	8.28
38	8–*f*–1	0.9	18.5	0.7	4000	5.83	0.175	5.71
39	11–*f*–1	1.7	14.6	0.4	4000	4.43	0.251	3.98
40	24–*k*–½	0.7	35.3	0.7	4000	12.6	0.294	3.40
41	10–*f*–1	1.5	20.6	0.4	4000	6.65	0.333	3.00
42	5–*d*–1	0.8	90.0	1.6	4000	19.9	0.531	1.88
43	13–g–1	4.2	115.0	6.4	4000	5.54	0.776	1.29
The Baird Bolometers								
44	slow	50	60	3.6	5000	1.00	1.67	0.600
45	fast	4.1	64.8	0.2	5000	16.0	2.19	0.457
The Thermistor Bolometer								
46	*A*	3.0	100	0.6	30	215	21.5	0.0465
The Polaroid Bolometer								
47	*A*	4.0	263	6.0	100	85	11.3	0.0884
The Aiken Bolometers								
48	*G* 165	3.8	185	27.5	1.85	210	26.6	0.0376
49	*G* 22	6.9	232	27.5	1.85	196	45.1	0.0222
The Golay Pneumatic Heat Detector								
50		5.0	0.48	7.0	1.0	1.28	(0.213)	(4.69)

were prepared by evaporating nickel upon a nitrocellulose film about 800A thick. The three reported in Table I contain four sensitive strips in the form of a cross. The measured zero-frequency responsivities of the three bolometers in the order specified in Table I are 118.5, 130, and 144.5 ohms/watt. Since the current was 15 milliamperes, the corresponding zero-frequency responsivities are 1.78, 1.95, and 2.17 volts/watt.

The frequency response curves of the three units are all of the type specified in Part I of Paper II as Class *I*. The frequency response curve of unit 324 has a slope of 3 db per octave at 34 c.p.s., at which point the response is 4.7 db less than its zero-frequency value. The reference time constant and the effective zero-frequency responsivity are thus 4.69 milliseconds and 1.46 volts/watt.

The frequency response curve of unit 347 has a slope of 3 db per octave at 22 c.p.s., at which point the response is 4.7 db less than its zero-frequency value. The reference time constant and the effective zero-frequency responsivity are thus 7.24 milliseconds and 1.60 volts/watt.

The frequency response curve of unit 350 has a slope of 3 db per octave at 44 c.p.s., at which point the response is 5.6 db less than its zero-frequency value. The reference time constant and the effective zero-frequency responsivity are thus 3.62 milliseconds and 1.61 volts/watt.

The Thermistor Bolometers (31–33) (N)

These bolometers were developed by Bell Telephone Laboratories under the direction of Dr. J. A. Becker. The sensitive element is a polycrystalline semi-conductor. Unit *S*-19 is glass-backed, unit *S*-20 is quartz-backed, and unit *XB*-108 is air-backed. The measured responsivities of the three units listed in Table I at a frequency of 15 c.p.s. are 675, 450, and 275 r.m.s. volts/r.m.s. watt.

The frequency response curves of units *S*-19 and *S*-20 are of the type designated in Part I of Paper II as Class *I*. The frequency response curve of unit *S*-19 has a slope of 3 db per octave at a frequency of 27 c.p.s., at which point the response is 2.3 db less than the response at 15 c.p.s. Accordingly, the reference time constant is 5.9 milliseconds and the effective zero-frequency responsivity is 0.7 db greater than its value at 15 c.p.s.: 730 volts/watt.

The frequency response curve of unit *S*-20 has a slope of 3 db per octave at 125 c.p.s., at which frequency the response is 4.5 db less than the response at 15 c.p.s. Accordingly, the reference time constant is 1.27 milliseconds and the effective zero-frequency responsivity is 1.5 db less than its 15 c.p.s. value: 378 volts/watt.

The frequency response curve of unit *XB*-108 corresponds to a single time constant of 135 milliseconds. Accordingly, the zero-frequency respon-

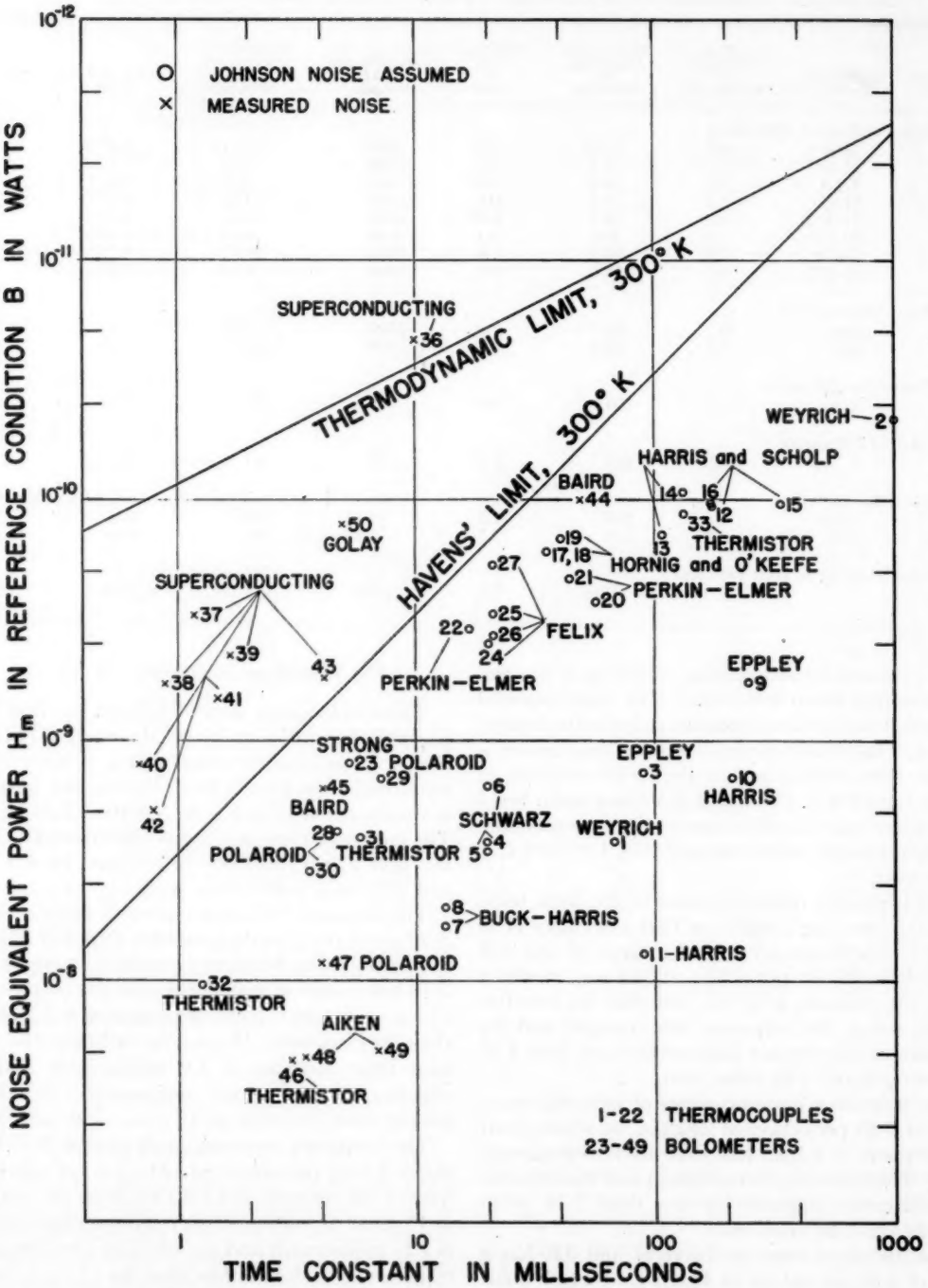

FIG. 1. Noise equivalent power in reference condition *B* as a function of the reference time constant for the approximately fifty Type *II* detectors listed in Tables I and II.

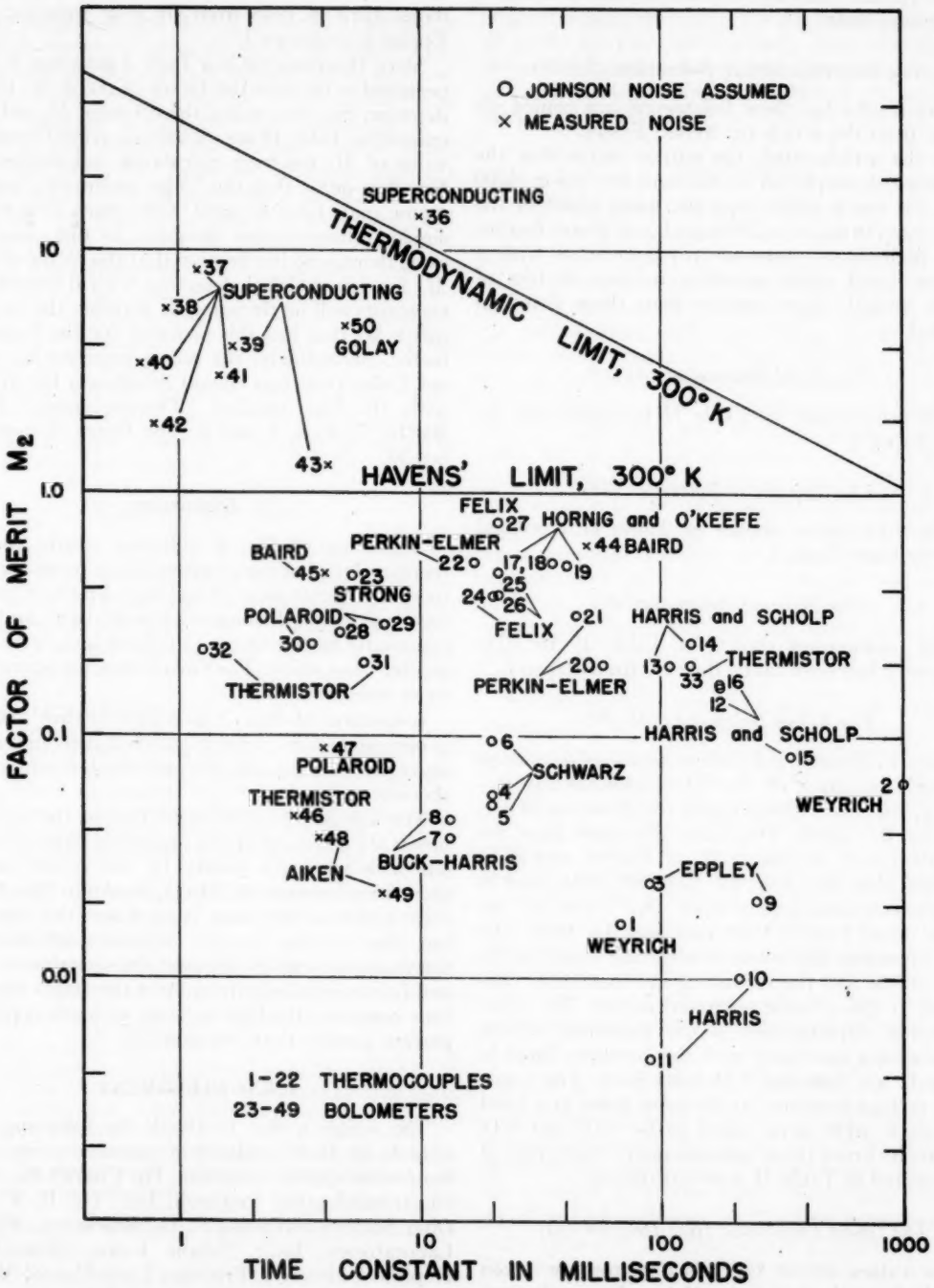

FIG. 2. Factor of merit M_2 as a function of reference time constant for the approximately fifty Type *II* detectors listed in Tables I and II.

sivity is 12.7 times the responsivity at 15 c.p.s., or 3500 volts/watt.

The Superconducting Bolometers (36–43)

The results for these bolometers are copied directly from the article by Nelson Fuson.[2]

In the article cited, the author states that the band width employed to measure the noise (4000 c.p.s.) is much wider than the band width of the detectors (16 c.p.s. to 230 c.p.s.), and states further that preliminary tests on two bolometers with a narrow band width amplifier indicate factors of merit several times smaller than those listed in Table II.

The Baird Bolometers (44–45)

The information in Table II is copied directly from Paper I.

The Thermistor Bolometer (46)

The information shown in Table II is copied directly from Paper I.

The Polaroid Bolometer (47)

The information shown in Table II for this bolometer has been taken directly from Paper I.

The Aiken Bolometers (48–49)

These evaporated gold bolometers were developed in the laboratory of the Electromechanical Research, Inc. by a group under the direction of Dr. Charles B. Aiken. The units discussed here are described in C. B. Aiken, W. H. Carter, and F. S. Phillips, Rev. Sci. Inst. **17**, 377–385 (1946) and in private communications from Dr. Aiken to the writer dated June 7, 1948, and June 11, 1948. The time constants and sensitive areas are stated in the 1946 article and the remaining information is contained in the private communications. The zero-frequency responsivities in the secondary of the transformers associated with the detectors listed in Table II are 8.66 and 7.44 volts/watt. The r.m.s. noise voltage measured at the same point in a band 1.85 c.p.s. wide were found to be 0.16 and 0.18 microvolt. From these measurements the values of P_0 reported in Table II were calculated.

The Golay Pneumatic Heat Detector (50)

The values stated for this detector are taken directly from the article by Nelson Fuson, J. Opt. Soc. Am. **38**, 845–853 (1948), who in turn obtained the information through a private communication dated April 26, 1948, from Mr. Roy Anderson of the Eppley Laboratory.†

Since this detector is a Type *I* detector, it is not permissible to state the factor of merit M_2 for this detector. For this reason, the values of M_2 and $1/M_2$ entered in Table II are enclosed in parentheses. The value of M_2 has been calculated and displayed in Fig. 2 in order that the Golay pneumatic heat detector may be compared with other detectors of similar reference time constant. In this case, however, there is no implication that the factor of merit M_2 for other Golay detectors with different time constants will be the same. It is rather the factor of merit M_1 that has this property for the Golay detector. Accordingly, the points representing different Golay detectors should lie along a line parallel with the line marked "Thermodynamic Limit, 300°K" in Figs. 1 and 2. The factor of merit M_1 is 0.30.

6. Discussion

Inspection of Fig. 2 indicates clearly that no thermopile or bolometer operating at room temperature has a factor of merit as large as unity. Only the superconducting bolometers (36–43) and the Golay pneumatic heat detector (50) have factors of merit greater than unity. The Golay detector operates at room temperature.

Inspection of Fig. 2 indicates further that the maximum factors of merit attained at room temperature are substantially the same for bolometers and thermocouples.

The suitability of the definition of the factor of merit M_2 proposed in the preceding paper is borne out with particular nicety by the results on the thermistor bolometers (31–33) shown in Fig. 2. The three units use the same type of sensitive element, but they employ greatly different coefficients of thermal conductivity between the sensitive element and its surroundings. In spite of the 100:1 range of time constant, the largest factor of merit is only 15 percent greater than the smallest.

ACKNOWLEDGMENT

The writer wishes to thank the following individuals for their kindness in communicating information on specific detectors: Dr. Charles B. Aiken, Electromechanical Research, Inc.; Dr. E. T. Bell, Ohio State University; F. G. Brockman, Phillips Laboratories, Inc.; Nelson Fuson, The Johns Hopkins University; Professor Louis Harris, Massachusetts Institute of Technology; William Langton, Baird Associates; and Max D. Liston, Perkin-Elmer Corporation.

[2] Nelson Fuson, J. Opt. Soc. Am. **38**, 845–853 (1948). The writer is indebted to Mr. Fuson for his kindness in sending a copy of his manuscript prior to its publication.

† See also Eplab Bulletin No. 10.

APPENDIX

The purpose of this section is to call attention to the wide variety of possibilities in the specification of the relevant properties of radiation detectors, and thereby to encourage precision in the statement of these properties in the periodical literature. In a letter to the writer from Nelson Fuson dated June 16, 1948, the suggestion is made that the present paper "will be particularly helpful to a lot of us who are working on one particular detector if you emphasize the need for stating *all* the experimental conditions, which we so often tend to overlook because they are, for our particular laboratory or test arrangement, kept constant with the result that we get careless and do not record them or think of them as variables."

Many of the matters mentioned below may seem too obvious to require specific mention, but they are included because of specific instances of failure to observe them in the literature.

Necessary Information

In the case of those detectors, such as the thermopile, for which it is reasonable and convenient to assume that the only source of noise is the Johnson noise associated with their resistance, it is sufficient to state the following information about the detector:

1. The electrical resistance.
2. The sensitive area.
3. A curve of relative responsivity *versus* frequency. If the frequency response may be characterized by a single time constant, a statement of its value is sufficient.
4. The relative response to different radiation wave-lengths.
5. The responsivity (in volts/watt) at a single frequency with a specified spectral energy distribution.

One may, of course, combine Items 3 and 5 by stating the responsivity in absolute terms as a function of frequency.

In the case of detectors for which it is not reasonable to assume that the only noise is Johnson noise, it is sufficient to state the following quantities:

1. The sensitive area.
2. A curve of relative responsivity *versus* frequency.
3. A curve of the relative noise power per unit band width *versus* frequency, under the same conditions used to determine Item 2.
4. The relative response to different radiation wave-lengths.
5. A single measurement of the signal-to-noise ratio under fully defined conditions.

Item 5 may be omitted if Items 2, 3, and 4 are on an absolute basis.

Sensitive Area

In stating the sensitive area of the detector in the case of thermocouples and thermopiles, it should be made clear just what area is involved. One may state the area of the receiver associated with a single junction, with all of the hot junctions, or with all of the hot *and* cold junctions. With that type of bolometer in which one strip is the load resistor of the other, one should state whether the specified area is that of one strip or of both strips.

Responsivity

The responsivity may be stated for a steady signal, for a sinusoidally modulated signal, for a square wave signal, or for isolated pulses of given shape. The magnitude of the radiation signal may be specified in terms of its peak-to-peak value, half of the peak-to-peak value, the average of the absolute value of the deviations from the mean (average value), the square root of the average of the square of the deviations from the mean (r.m.s. value), and each of these may refer to the complete signal, or only to the Fourier component with the fundamental frequency. The same comments apply to the specification of the resultant voltage response.

Thus the statement that the responsivity at a certain frequency is so many volts per watt contains considerable ambiguity; for example, the responsivity in peak-to-peak volts per r.m.s. watt is eight times the responsivity in r.m.s. volts per peak-to-peak watt, when the radiation is sinusoidally modulated.

Furthermore, it should be specified whether the responsivity is that corresponding to open circuit conditions, or if not, just what the impedance relations are. This problem is particularly acute with bolometers, because it is necessary to feed current through the bolometer from a load resistor; unless the load resistor has a resistance large compared with the bolometer resistance, it reduces the factor of merit for the bolometer.

The same uncertainties of specification that affect the sensitive area are also of concern here, since the responsivity in volts per watt is usually obtained from the responsivity in volts per watt per unit area, by multiplying the latter by the sensitive area.

Frequency Response

Even the curve of relative frequency response is not independent of the specification of the measurement of the voltage and of the radiation. Consider, for example, a detector with a single time constant equal to 15.9 milliseconds. Then the response of the detector is 3 db less than its low frequency value at 10 c.p.s. as measured by the r.m.s. response to a sinusoidal radiation signal, but the response as measured by the peak-to-peak response to square wave irradiation will not be 3 db less than its zero frequency value until the frequency rises to 17.8 c.p.s.

Thus even in specifying the relative frequency response curve, it is necessary to be careful to describe adequately the method of measurement.

Provided that the relative frequency response curve cannot be specified by a single time constant, it is desirable to exhibit the actual curve, preferably with logarithmic coordinates. If, however, for reasons of brevity this is not possible, then the minimum essentials are the specification of the frequency at which the response is down 3 db, and the specification of the frequency and attenuation factor for that point on the curve at which the slope is 3 decibels per octave, when the response is measured with sinusoidally modulated radiation.

Resistance

A specified resistance may refer to the resistance of a single junction, to the resistance of all of the hot junctions, or to the resistance of all of the hot *and* cold junctions.

With bolometers, the resistance may be that of the sensitive strip, or it may be that of the strip plus a load resistor; in some cases the latter may be another strip.

Noise Level

When measurements of the signal-to-noise ratio or the noise level of the detector are specified, it is essential to state the noise equivalent band width employed, and to state what property of the noise is measured: One may state the peak-to-peak, half of the peak-to-peak, average, or r.m.s. voltage of the noise.

In general, one should avoid specifying the noise level in terms of the peak value as observed on an oscilloscope, because a peak of any preassigned value may be observed if one waits long enough. Even if one is careful to measure the average maximum occurring in specified time periods, the relationship[3] of such measurements to the more direct measurements of the r.m.s. or average value is not a simple one, and depends on the power spectrum of the noise.

It is furthermore necessary to state the noise level at the same point in the circuit as that at which the responsivity is stated. For example, when a coupling transformer is used, it is not adequate to give the noise at the input grid of the amplifier, and the responsivity of the detector itself, unless further information is provided about the properties of the transformer.

Spectral Response

It has been customary to assume that the receiving area of the detector is equally black at all wave-lengths, so that the only relevant property of the radiation falling upon the detector is the total power incident upon the receiving area.

As the techniques of both construction and measurement improve, however, the error involved in this assumption will no longer be tolerable, and it will be necessary to specify the spectral response of the detector. Furthermore, in specifying the noise equivalent power in the reference condition, it will be necessary to specify the spectrum of the source used.

In the case of photo-conductive cells, this necessity has been recognized from the beginning, because of the large variation of sensitivity as a function of wave-length.

[3] For a detailed treatment of this relationship, see S. O. Rice, Bell Sys. Tech. J. **24**, 46–156 (1945), especially p. 75.

Erratum: Reflected Light and Ghosts in Optical Systems

[J. Opt. Soc. Am. 39, 30 (1949)]

ALLEN E. MURRAY
Scientific Bureau, Bausch and Lomb Optical Company, Rochester, New York

ON page 33 the left-hand column should read as follows:

first reflection at last surface
first reflection at penultimate surface
. . . . . .
first reflection at third surface
first reflection at second surface

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 5 MAY, 1949

The Spectra of the Heavy Elements

FRANK S. TOMKINS
Argonne National Laboratory, Chicago, Illinois

AND

MARK FRED
Argonne National Laboratory and Armour Research Foundation, Chicago, Illinois

(Received February 4, 1949)

The spectra of the new heavy elements are characterized by great complexity similar to that well known for thorium and uranium. Consequently descriptions of the spectra to be adequate for term analyses and assignment of electron configurations must be quite extensive, containing complete, accurate wave-lengths and much other material. Since considerable effort and amounts of sample are required it will be some time before such descriptions are available. On the other hand it would be useful for spectrochemical analyses to have wave-lengths to 0.1A and intensities in a convenient spectral region, and this information has been collected for protactinium, neptunium, plutonium, and americium. Intensities for each line were obtained as the reciprocal of the limiting dilution at which it could be seen with the copper spark source, and therefore the same intensity scale is used for each element. Hence it could be observed that the spectrum of americium appears to stand in the same relation to those of the other heavy elements that the spectrum of europium does to those of the preceding rare earths, both elements containing relatively few lines of high average intensity. This suggests by analogy that Am *I* contains seven *f* electrons plus two others.

THIS paper lists the most prominent spectrum lines of the elements from protactinium to americium. In most cases the spectra were made available because of the necessity of conducting spectrochemical analyses on a number of samples for impurities.* Because of limitations of sample size it was not feasible to perform the analyses with the aid of chemical or physical[1] separations and so the spectrum of the major constituent was always present. It would be of considerable interest to provide descriptions of the spectra which would be adequate for term analyses leading to energy levels and electron configurations, not only for the spectroscopic interest but also for the assistance they might provide in the correlation of chemical properties of these elements. However, the amounts of labor and sample required for such descriptions are quite formidable, as is evidenced by the fact that the spectra of thorium and uranium are still incompletely analyzed in spite of having been available in unlimited quantities for many years. To be adequate for a term analysis the wave-lengths must be measured to the third decimal place for as many lines as can be observed over the most extensive wave-length range accessible; Zeeman patterns must be provided; and a temperature or excitation variation provided in order to distinguish the various stages of ionization and low levels from high levels. On the other hand, for purposes of identification it is sufficient to have measurements of the strongest lines of each element, accurate to the first decimal place, covering a representative wave-length region which is convenient for analysis. Since it will probably be some time before data of much value for term analyses are accumulated, the following information was collected for whatever usefulness it might have in spectrochemical analyses. However, it was found on closer examination of the assembled data that statistical considerations indicated that the analogy between the heavy elements and the rare earths was closer than had been anticipated, and in fact was sufficient to permit some conclusions to be made regarding electron configurations.

The spectra were taken on a Baird three-meter spectrograph which covers a range of 1400A in the first order. For comparison with the standard plates of the elements which had been prepared for analytical purposes, the samples were exposed in the region 2650–4050A, corresponding to an angle of incidence of 3.3°. A second exposure was usually made at an angle of incidence of 11.5°, covering the region 3700–4370A in the second order, since this region is quite useful for the detection of many elements with complex spectra such as the rare earths. The intensity distribution of the particular grating employed in the spectrograph is such that equal background intensities are obtained for the same exposures at the two angles of incidence, and consequently the intensities observed in the second order were divided by two to reduce all intensities to the first order scale. The wave-lengths were measured by comparison with neighboring iron

* The spectrum of plutonium was first photographed by Rollefson and Dodgen, of the University of California at Berkeley, who reported [G. K. Rollefson and H. W. Dodgen, Plutonium Project Report CK-812, July, 1943] wave-lengths and arbitrary intensities for about 125 lines, later extended [H. W. Dodgen, J. Chrisney, and G. K. Rollefson, Plutonium Project Report CK-2247, October 9, 1944] to about 700 lines.

[1] B. F. Scribner and H. R. Mullin, J. Research Nat. Bur. Stand. **37**, 379 (1946).

TABLE I. Sensitive lines of protactinium.

Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks
2640.3	100		3033.6	50		3705.9	10		3970.0	50	
2644.9	20		3034.3	50		3713.4	10		3975.4	20	
2651.6	20		3036.7	20		3716.8	20		3979.5	10	
2670.5	50		3039.0	20		3717.3	10		3980.1	10	
2672.1	50		3042.0	100		3725.0	10		3980.3	10	
2697.5	50		3042.7	50		3725.6	10		3982.2	20	
2732.2	100		3045.6	50		3729.3	20		3983.3	20	
2743.2	100		3051.8	20	VW	3730.4	10		3988.3	20	
2743.9	200		3053.5	100		3731.0	10		3997.7	10	
2755.9	20		3054.6	100		3735.7	20		3999.4	10	
2793.3	50		3057.9	20		3737.8	20		4000.2	10	
2796.2	100		3058.1	10		3743.5	10		4000.9	10	
2808.0	20		3059.1	20		3754.3	20	VW	4009.4	10	
2811.5	50		3060.0	10		3763.5	10		4012.4	10	
2815.4	20		3060.3	20		3764.1	10		4012.9	50	VW
2826.6	50		3060.9	20		3765.7	20		4018.2	50	VW
2832.6	50		3066.4	50		3780.9	20		4029.9	20	VW
2835.5	50		3067.7	20		3781.3	10		4046.9	20	
2843.3	50		3068.4	10		3793.7	10		4047.6	20	
2845.5	50		3069.1	10		3795.6	20		4051.8	20	
2855.4	20		3070.3	20		3797.7	10		4056.1	50	VW
2870.0	50		3071.2	10		3798.0	10		4070.4	50	VW
2871.4	50		3075.6	20		3802.8	20		4074.3	20	
2873.3	50		3076.7	20		3803.3	10	VW	4087.8	20	VW
2891.0	20		3081.3	10		3807.4	20		4089.4	20	VW
2893.0	100		3082.3	20		3814.0	20		4090.4	20	
2906.3	20		3083.2	50		3818.2	20		4099.2	20	
2906.9	50		3092.6	20		3823.6	20		4102.1	50	VW
2907.5	50		3093.2	50		3828.4	20		4107.9	20	
2909.0	50		3103.8	20		3835.4	10		4108.4	20	
2909.6	100		3105.9	50		3849.9	50	VW	4110.8	10	
2922.8	50		3111.7	50		3856.9	50	VW	4117.1	10	
2923.4	10		3117.7	20		3865.4	10		4128.7	20	
2928.3	50		3131.2	20		3866.8	10		4129.3	20	
2932.9	20		3136.9	20		3868.5	20		4139.6	20	
2934.3	50		3157.6	20		3875.9	50		4153.4	10	
2940.2	100		3170.8	20		3878.5	10	VW	4159.2	10	
2943.0	50		3171.5	20		3890.6	10		4160.1	50	
2943.5	100		3175.9	10		3891.5	10		4161.0	20	
2954.6	50		3178.2	20	VW	3892.6	20		4161.5	20	
2959.7	100		3179.6	20	VW	3905.0	20		4164.2	50	
2968.0	50		3203.1	10		3908.0	50	VW	4176.1	20	VW
2969.0	50		3204.2	20		3909.1	20		4189.0	10	
2970.7	20		3205.4	20		3909.6	20		4192.1	50	
2972.4	20		3205.9	10		3911.3	20		4196.9	50	
2974.0	50		3214.0	50		3915.1	10		4197.8	50	
2977.2	20		3221.9	20		3916.2	10		4205.3	20	
2980.5	100		3240.5	20		3921.5	20		4210.4	20	
2981.9	20	VW	3332.8	20		3923.9	10		4210.9	10	
2982.9	20		3394.4	10		3926.0	10		4217.2	100	
2983.4	20		3448.4	10		3926.9	10		4225.0	10	
2984.7	50		3452.8	10		3928.7	50		4230.6	50	
2985.3	10		3468.2	10		3944.0	20		4231.0	10	
2986.4	10	VW	3471.1	10		3945.9	20	VW	4248.1	100	
2987.0	20		3480.8	10		3952.6	20		4286.7	50	
2987.9	100		3488.9	10		3953.4	10		4291.3	50	
2991.6	20	VW	3545.0	50		3956.6	20		4298.4	10	
2992.7	50		3632.4	50		3957.8	100	VW	4299.4	20	
2994.9	50		3643.8	20		3959.1	10		4304.0	10	VW
2997.7	50		3663.2	20		3961.5	50		4311.3	10	
3004.6	100		3665.6	10		3961.8	10		4329.2	10	
3005.7	50		3666.3	10		3962.3	10		4338.8	10	
3009.5	50		3668.4	10		3962.9	10		4361.9	10	
3015.2	50		3674.9	20		3964.2	10		4369.3	10	
3023.5	50		3679.8	20		3965.7	10		4371.7	20	
3025.3	100		3692.8	20		3967.3	20				

lines using a 0.1-mm scale, taken from a Bausch and Lomb measuring eyepiece, in an Applied Research Laboratory comparator-densitometer which provided dual projection of spectra and scale on a ground glass screen with a twentyfold magnification. The wave-lengths should be accurate to better than 0.1A.

The spectra were obtained by the copper spark

TABLE II. Sensitive lines of neptunium.

Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks
2655.0	20		3057.5	20		3870.9	10	W	4156.3	10	W
2669.6	20		3070.3	20	W	3883.1	10	W	4164.5	50	
2678.2	20		3078.0	50		3888.2	20	W	4165.3	10	
2733.7	20		3084.7	20		3929.3	10		4172.0	10	VW
2734.5	20		3090.4	20		3932.2	10		4182.1	10	W
2785.2	20		3092.0	20		3949.2	20		4192.7	10	
2789.8	20		3092.5	20		3949.7	10		4196.6	10	
2833.6	20		3164.8	10		3951.0	10		4197.1	10	
2841.0	20		3171.6	20		3954.4	10		4208.5	10	
2841.5	20		3185.9	10		3961.6	10		4209.7	10	
2864.1	20		3315.7	10		3967.4	10	W	4234.8	10	
2864.4	20		3337.0	10		3987.0	20		4246.8	10	
2865.5	20		3402.5	20		3988.6	20		4256.7	20	
2866.0	20		3589.2	10		3989.8	20		4258.1	20	W
2866.9	10		3590.6	20		4000.4	10	W	4269.6	10	W
2867.6	20		3665.6	20	W	4004.2	10		4273.2	10	
2869.8	20	VW	3708.3	20	W	4015.5	10		4279.6	20	W
2873.3	20		3711.8	10	W	4028.9	20	W	4281.4	10	W
2888.5	10		3716.2	10	W	4031.6	10		4289.7	10	
2956.6	50		3722.3	20	W	4031.8	10		4290.9	50	
2957.9	50		3730.7	10		4050.1	10	VW	4296.1	10	W
2968.2	20		3751.0	10		4051.1	10	W	4307.8	20	
2971.1	20		3780.6	10		4077.4	10		4319.8	10	W
2974.3	50		3792.9	10		4086.8	10		4333.9	10	
2974.9	50		3794.9	20		4098.8	10	W	4336.6	20	
3026.4	50		3829.2	50	VW	4108.4	20	W	4351.9	10	W
3027.3	20		3832.3	20	W	4121.2	10		4363.8	10	
3029.1	20		3853.2	10	W	4123.3	10				
3052.0	50	W	3865.2	20		4136.9	10				

method[2] developed in this laboratory for such applications. This method has the advantage, in addition to high absolute sensitivity, of providing fairly reproducible intensities. Hence a measure of the intensity of each line could be obtained conveniently by exposing successively diluted samples and determining the limiting amount of the element at which the line could be distinguished above the continuous background of the source. These limits were obtained by visual inspection and should be accurate to a factor of two. The intensities were then expressed as one hundred times the reciprocal of the limiting weight in micrograms on the electrode, the factor being chosen to avoid fractional

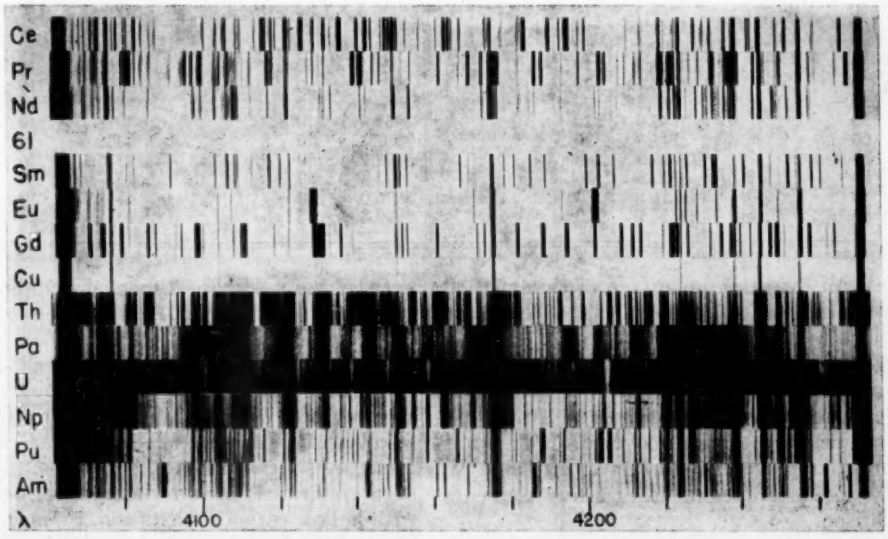

FIG. 1.

[2] Fred, Nachtrieb, and Tomkins, J. Opt. Soc. Am. **37**, 279 (1947).

TABLE III. Sensitive lines of plutonium.

Wavelength	Intensity	Remarks	Wavelength	Intensity	Remarks	Wavelength	Intensity	Remarks	Wavelength	Intensity	Remarks
2677.0	20		2981.2	20		3743.1	10		3980.4	20	
2682.9	10		2986.9	20		3753.6	20		3984.0	10	
2683.4	10		2988.1	100		3755.9	20		3985.5	100	
2684.8	20		2991.5	20		3758.3	20		3988.5	20	
2693.3	20		2993.9	100		3772.8	10		3989.7	100	
2781.3	50		2996.4	100		3773.6	20		3990.2	10	
2784.4	50		3000.4	100		3774.4	10		3991.5	20	
2787.2	20		3008.9	50		3775.7	20		3992.2	20	
2803.9	10		3009.6	50		3776.8	10		4010.6	10	
2806.0	20		3028.8	20		3792.2	20		4013.1	10	
2808.4	50		3042.5	20		3803.6	20		4015.9	20	
2809.0	20		3043.0	20		3810.2	20		4021.5	50	
2815.6	50		3060.2	10		3812.3	20		4048.9	10	
2826.2	20		3069.3	20		3814.9	20		4064.6	20	
2833.2	50		3089.6	10		3823.9	20		4066.8	20	
2835.5	100		3090.5	10		3827.5	20		4078.0	20	diffuse
2898.0	50	diffuse	3091.2	20		3835.5	20		4101.0	20	
2899.8	20		3091.9	20		3836.9	20		4102.0	20	
2904.3	20		3092.7	50		3842.1	10		4104.4	10	
2905.0	50		3093.3	20		3851.9	10		4105.9	10	
2910.3	50		3104.1	20		3852.7	50		4107.2	10	
2912.7	20		3105.0	20		3864.4	10		4116.4	20	
2913.7	20		3105.9	20		3864.7	10		4140.0	10	
2914.2	20		3117.8	10		3865.0	10		4141.2	20	
2914.9	20		3118.5	10		3869.7	10		4190.0	10	diffuse
2915.6	50		3123.8	20		3870.1	10		4196.2	50	
2918.0	20		3136.5	20		3874.2	20		4204.7	10	
2925.2	100		3159.2	50		3878.6	20		4205.5	10	
2925.8	50		3161.5	10		3886.4	10		4206.4	10	
2926.3	20		3163.1	10		3886.5	10		4208.0	10	
2928.2	50		3174.4	10		3887.4	20		4208.2	10	
2929.5	10		3179.3	20		3892.7	20		4208.8	10	
2930.9	20		3198.3	10		3895.9	20		4219.0	10	
2932.4	20		3220.8	10		3904.1	50		4229.7	20	
2933.2	20		3221.4	20		3907.2	100		4247.2	10	
2933.8	20		3231.8	20		3910.3	20		4254.4	20	Cr 4254.35?
2937.6	50		3232.6	20		3912.5	50		4254.8	20	
2938.4	10		3312.6	20		3913.4	50		4256.0	20	
2938.9	100		3401.0	20		3928.0	20		4257.9	20	
2941.3	50		3465.0	10		3928.5	20		4261.8	10	
2945.2	50		3469.2	10		3931.2	20		4273.3	100	
2945.9	50		3473.6	10		3944.0	10		4278.4	10	
2950.0	20		3585.9	20		3947.0	10		4278.8	10	
2951.6	50	VW	3627.5	10		3949.2	20		4280.3	20	
2954.3	100		3632.1	20		3953.0	20		4283.7	10	
2964.5	20		3709.1	20		3958.9	20		4289.1	20	diffuse
2966.7	20		3714.7	10		3961.5	50		4297.8	10	
2967.3	20		3718.1	20		3962.8	20		4298.2	10	
2968.9	50		3720.2	20		3963.6	20		4314.3	10	
2970.2	50		3720.6	20		3964.9	10		4316.7	10	
2972.3	100		3721.5	20		3967.3	20		4330.6	20	
2977.1	20		3726.0	50		3972.2	50		4336.1	10	
2977.8	50		3726.8	50		3975.4	50		4337.2	20	
2978.4	20		3732.1	20		3976.0	20		4352.7	50	
2980.0	50		3739.4	10		3979.8	10		4358.1	20	

intensities. An essential advantage of this method of indicating intensities is the fact that all elements are described by the same intensity scale.

The spectra of protactinium, neptunium, plutonium, and americium are given in Tables I, II, III, and IV, respectively. The most sensitive and distinctive lines in the region covered for the kind of excitation employed are collected in Table V, which also includes lines of other heavy elements for comparison.

In the spectra of protactinium and neptunium the existence of wide hyperfine structure was apparent even at the moderate dispersions employed (5.5A/mm in the first order, 2.7A/mm in the second order). One line of neptunium in particular, 3829.2, appeared to be very wide on a second-order plate and an exposure was made in the third order of the spectrograph with a 12.5-microgram sample which resolved the first three lines of a flag pattern with probably six components. This structure has since been confirmed[3] by interferometric measurements with hollow cathode excitation, by means of which approximately fifty other neptunium lines were

[3] Frank S. Tomkins, Phys. Rev. **73**, 1214 (1948).

TABLE IV. Sensitive lines of americium.

Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks	Wave-length	Intensity	Remarks
2661.6	20		3037.8	100		3753.2	100		4107.0	20	
2664.8	20		3047.6	20		3757.7	200		4108.8	20	
2668.8	20		3057.4	20		3761.6	500		4109.5	20	
2686.9	20		3059.1	50		3767.1	20		4109.9	20	
2691.4	100		3071.3	100		3769.0	20		4124.5	20	
2706.4	50		3082.8	100		3774.3	20		4127.4	20	
2716.5	50		3162.1	500		3777.4	500		4128.3	20	
2720.5	20		3162.9	100		3788.7	50		4128.9	20	
2725.4	20		3203.3	100		3789.0	20		4130.7	20	
2727.6	20		3314.0	100		3801.7	100		4136.7	20	
2728.7	200		3332.1	100		3848.9	100		4137.7	20	
2731.9	20		3362.5	50		3871.5	500		4144.0	20	
2732.1	20		3392.0	50		3892.9	20		4148.0	20	
2735.4	100		3428.0	20		3896.0	20		4149.9	20	
2739.6	100		3431.6	20		3901.5	50		4154.7	50	
2742.3	50		3439.8	200		3904.2	100		4182.6	20	
2743.2	100		3446.0	50		3914.3	100		4184.0	20	
2746.6	100		3448.2	50		3916.0	500		4184.3	20	
2747.0	100		3452.0	100		3921.5	500		4186.6	50	
2749.4	500		3509.5	50		3926.2	1000		4188.2	500	
2749.8	20		3510.2	100		3936.2	100		4188.8	20	
2753.4	20		3547.9	100		3942.7	20		4189.6	20	
2755.3	200		3559.6	50		3943.8	50		4197.6	20	
2755.9	200		3561.8	50		3944.0	100		4198.7	20	
2756.7	500		3562.6	200		3944.5	50		4200.0	50	
2776.0	50		3569.2	100		3945.2	20		4219.4	20	
2776.5	50		3570.0	100		3946.0	50		4221.0	20	
2796.8	50		3573.1	100		3948.0	20		4226.8	100	
2812.9	200		3573.6	100		3950.3	20		4233.6	20	
2815.3	100		3581.2	200		3952.5	500		4234.0	20	
2816.0	50		3584.7	100		3963.3	50		4248.4	500	
2816.4	50		3593.7	100		3973.1	100		4249.4	20	
2832.3	2000		3615.8	20		3975.2	50		4249.9	50	
2852.2	100		3616.5	20		3982.6	20		4252.0	20	
2861.6	50		3617.4	20		3984.6	50		4261.8	50	
2873.3	100		3631.4	50		3991.9	50		4262.4	20	
2881.6	50		3639.5	20		4000.6	50		4265.6	200	
2888.5	200		3673.1	200		4003.0	20		4272.2	20	
2893.3	100		3683.4	50		4006.2	20		4278.2	20	
2899.6	100		3696.4	200		4009.2	100		4289.3	100	
2911.2	500		3706.2	100		4013.7	100		4289.9	20	
2920.6	500		3707.9	100		4017.9	20		4300.9	20	
2928.0	100		3710.3	100		4019.7	100		4302.5	20	
2937.0	100		3712.9	50		4020.2	50		4302.8	20	
2957.0	200		3720.0	50		4021.3	20		4309.7	100	
2966.8	500		3720.5	50		4027.4	50		4309.9	50	
2969.4	500		3721.2	20		4027.8	50		4324.6	500	
2972.6	200		3725.4	50		4032.9	20		4332.8	100	
2982.2	100		3733.0	20		4033.5	20		4335.6	50	
2993.5	50		3734.7	20		4036.4	200		4337.2	20	
2994.4	20		3737.1	20		4057.8	50		4341.8	20	
3004.3	200		3737.6	50		4069.8	50		4348.3	50	
3020.7	100		3740.4	50		4079.7	20		4349.3	50	
3021.1	50		3748.3	20		4080.5	20		4366.7	100	
3028.0	200		3749.4	50		4089.3	1000		4372.4	50	
3028.8	20		3751.0	50		4099.0	20		4374.9	50	
3037.4	20		3752.6	100		4102.3	20				

found which also had six components, indicating that the nuclear spin of Np^{237} is 5/2. Hyperfine structure in protactinium has been known for some time.[4] In Tables I and II the lines exhibiting unresolved hyperfine structure are designated by the symbol "W" or "VW."

Comparison of the tables shows that all the elements through plutonium have spectra which are quite complex, containing many lines of low and comparatively uniform intensity, with little to distinguish each element except details of wave-length. By contrast the spectrum of americium is comparatively simple, containing few lines of much greater intensity than the other elements. To make this comparison more significant a representative region of the spectrum was examined in detail for all of the elements and compared with rare earth elements. The region is defined by two strong copper lines at 4062.6 and 4275.1, respectively, and is shown in Fig. 1. The figure contains exposures of ten microgram amounts of the individual rare

[4] H. Schuler and H. Gollnow, Naturwiss. 22, 511 (1934).

TABLE V. Sensitive lines of the heavy elements.

Element	Wave-length	Intensity
Th	4019.1	1000
Pa	3957.8	100
	3054.6	100
	3053.5	100
U	3932.0	100
Np	4290.9	50
	3829.2	50
Pu	3989.7	100
	3907.1	100
Am	2969.4	500
	2832.3	2000

earths and 100 microgram amounts of the heavy elements with a copper blank in between, exposed in the second order of the spectrograph. The tenfold increase in sample size for the heavy elements corresponds to the intensity ratio of the strongest lines of the average rare earth and heavy element, but it is clear from the complexity of the heavy element exposures shown in the figure that the latter contain many more weak lines. The much greater intensity of the strong europium and americium lines is difficult to reproduce in a figure of this sort. In order to make this comparison more evident the lines in this region have been tabulated by intensity groups for each element and are presented in Table VI. (Some weak lines of intensity 10 included in Table VI were omitted from Tables I, II, and III.) By dividing the total intensity listed for each element by the number of lines an average line intensity for each element can be obtained. These average intensities are plotted in Fig. 2. The

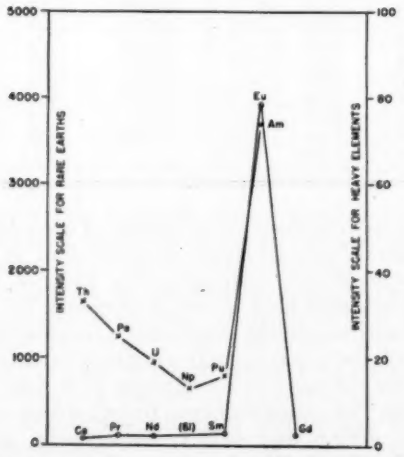

FIG. 2. Average line intensity for rare earths and heavy elements in the region 4062–4275A.

comparatively uniform averages shown for the first five members of each group and the much greater average intensity of europium and americium suggest that there is a close analogy between the two groups of elements.

The relatively simple spectrum of europium is due to the fact that it contains seven f electrons which form half of a closed shell of f electrons. It is a characteristic of half-closed shells that, while they produce a great many terms in the energy level diagram, there is one term of maximum possible multiplicity and zero orbital angular momentum which is considerably lower in energy than all the other terms of the configuration, the separation being of the order of 15,000 or 20,000 cm^{-1}. In the case of the configuration f^7, the lowest term is $^8S_{7/2}$. Because of the considerable energy difference between the lowest term and the others, the latter are practically never populated. Hence we may say that the situation in the ground state of the neutral europium atom in which two s electrons are added to the f^7 core of Eu *III* is analogous to the situation in neutral barium in which two s electrons are added to the xenon-like core of Ba *III*. The only difference for practical purposes is that in the former case the parent term of the core is an octet instead of $^1S_0$, the increased multiplicity leading to a comparatively small increase in the number of terms of neutral and singly ionized europium, compared with barium. On the other hand, a configuration consisting of other than seven f electrons, as in samarium, produces terms which, while slightly less numerous, are much more closely spaced and well populated. Hence the spectra of these elements are roughly equivalent to that of europium repeated a very large number of times. The effective number of such levels is not greatly different for different numbers of f electrons in the core, which explains the relatively constant line intensities of different rare earths. In the present case the important distinction between the different rare earths is the fact that the Eu *III* core consists of only one term for practical purposes. These facts are illustrated in Fig. 3, which compares the gross features of the energy level diagrams of barium and europium.

On the basis of these considerations, the simple spectrum of americium is strong evidence for the existence of the presence of seven f electrons. It may be noted that this finding is not in agreement with the prediction of W. F. Meggers[5] who assigned to the neutral americium atom the configuration $5f^6 6d 7s^2$. This prediction apparently was due in part to the fact that the neutral uranium atom has the configuration $5f^3 6d 7s^2$ instead of having four $5f$ electrons and no d electron, which one might ex-

[5] W. F. Meggers, Science 105, 514 (1947); W. M. Welch Manufacturing Company, "Periodic Chart of the Atoms," 1947 Edition.

TABLE VI. Statistical comparison of rare earth and heavy element spectra, 4062–4275A.

Element	No. lines in each intensity group 10	20	50	100	200	500	1000	2000	5000	10000	20000	Total intensity	No. lines	Intensity per line
Ba	0	1	0	0	0	1						520	2	260
La	0	0	1	0	4	3	1	3	1	1		24350	14	1739
Ce	20	29	19	32	6	3						7630	109	70
Pr	8	11	18	9	6	4	1					6300	57	111
Nd	5	13	18	8	8	3						5500	55	100
61														
Sm	5	9	17	18	11	5	1					8850	66	134
Eu	0	0	3	1	1	0	1	0	0	1	1	31450	8	3931
Gd	3	6	12	9	9	1	1					4950	41	121
Ra	0	0	1	0	0	0	0	1				2050	2	1025
Ac														
Th	26	27	10	4	3							2300	70	33
Pa	24	15	8	2								1140	49	23
U	43	20	7	0	1							1380	71	19
Np	20	3	1									310	24	13
Pu	28	12	1	1								670	42	16
Am	0	33	6	1	1	2	1					3260	44	74

pect by analogy with neodymium. Since we believe that the normal state of Am *I* is $5f^7 7_2^2$, by analogy with europium, it is uncertain just where the correspondence between the two groups of rare earths breaks down in between uranium and americium. It also seems probable from the rare earth analogy that the spectrum of curium will be complex and resemble that of gadolinium. These deductions regarding electron configurations are in agreement with the chemical and physical properties of these elements and their compounds.[6]

The fact that the americium spectrum is comparitively simple has led to the hope that it might be possible, with comparatively little effort, to achieve a successful term analysis by inspection using as a guide the Eu *I* and Eu *II* spectra which have already been analyzed. Examination of the europium spectrum as excited in the copper spark source used in this study shows that practically all the lines observed belong to Eu *II* which is, from the considerations enumerated above, almost alkali-like. From the fairly close correspondence of the energies of the $6p$ and $5d$ electrons in Eu *II* with those in Ba *II*, it was expected that the locations of these configurations in Am *II* could be judged from the corresponding configurations in Ra *II*, which would displace corresponding multiplets slightly toward the violet and increase the multiplet separations two- to threefold, compared to europium. Unfortunately, the similarity is not readily apparent. It is not believed that this circumstance casts doubt on the interpretation of the electron configurations of americium but is probably due to the larger multiplet separations which produce more overlapping. It is believed that a beginning on the term analysis of Am *II* can nevertheless be made without great difficulty, and to this end the description of the americium spectrum is being improved.

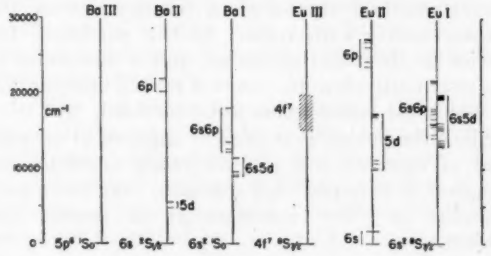

FIG. 3. Low terms of barium and europium.

[6] G. T. Seaborg, "Electronic structure of the heaviest elements," Plutonium Project Record, **14B**, Paper 21.1 (to be published).

A Note on the Photo-Pupil Reflex

JULES DE LAUNAY*
Naval Research Laboratory, Washington, D. C.
(Received January 19, 1949)

A study of the variation of pupil diameter with illumination of the retina indicates that the photo-pupil reflex is confined to the photopic range of illumination. On the basis of this result, the hypothesis is proposed that the photo-pupil reflex is essentially initiated by cone response. Suggestions are made for further investigation.

INTRODUCTION

THE purpose of this note is to present the preliminary results of the author's program of study of the photo-pupil reflex and to discuss some possible directions of further investigation which may prove to be fruitful.

The note includes a description of the experimental method employed, a presentation of the measurements with regard to the statistical behavior of the pupil diameter, and a discussion of the essentially photopic range of retinal illumination to which the reflex seems to be confined.

Since the author is at present engaged in another field of research and can no longer continue the program, it is hoped that this note may serve as a stimulus to other investigators to pursue the subject.

EXPERIMENTAL METHOD

The object of the experimental arrangement was to produce a specified stimulus on the retina and to photograph the pupil after it had had sufficient time to react to this stimulus.[1] The stimulus consisted of a 2° circular disk of white light falling on the fovea, or in the periphery (8° temporally from the fovea centralis), as desired. The value of 2° was selected for stimulus size because this value was a little less than the diameter of the author's rod free foveal region. The stimulus was obtained by viewing through an artificial pupil of one square millimeter area located at the anterior focus of the eye, a magnesium oxide-coated target subtending 2° at the eye. Both eyes were open, but only the left eye was exposed to the stimulus. The target was an ellipse set at an angle of 45° with the direction of vision, with the major and minor axes chosen so as to produce a 2° circular image when viewed at a distance of one meter. The unit of brightness chosen was the meter-candle-millimeter (mcm) which is defined for practical reasons as the brightness of a target freshly coated with magnesium oxide smoke, illuminated at a distance of one meter by a source of one candlepower, and viewed through an artificial pupil of one-square-millimeter opening, placed at the anterior focus of the eye. This definition of stimulus brightness ignores personal factors such as the absorption by the optical media of the eye. Fortunately, this is not too important for this particular study.

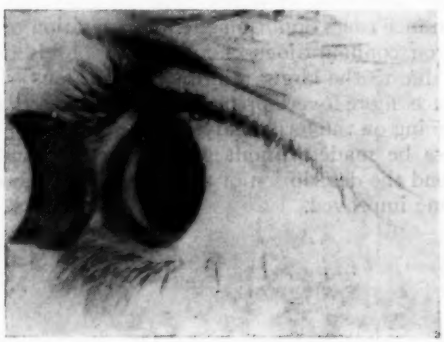

FIG. 1. The eye in the absence of retinal illumination. The flash picture is made before the eye can blink or the pupil constrict.

The eye was dark adapted before each exposure to the stimulus. It was then exposed to the stimulus for some ten or more seconds[1] and a photograph taken. The camera positioned at about 45° from the line of sight and placed 6 inches from the eye, was set for time exposure and the eye illuminated by a flash bulb from a temporal direction of about 70°. Though the first flash or two in a series of observations produced a positive after-image and considerable light adaptation, later flashes fell on the temporary blind spot (negative after-image) caused by the first flash. Thus, for weak illumination of the target the time required for dark adaptation between successive photographs and removal of the positive after-image of the flash bulb was reduced to about 15 minutes. It was assumed that provided the positive after-image in the 70° nasal region of the retina had disappeared and the rest of the retina dark adapted (as shown by tests), the temporary scotoma at 70° had no effect on the photo-pupil reflex.

* This research was performed in the Physics Department of Texas Technological College during the years 1942–1946.

[1] Prentice Reeves, Psych. Rev. 25, 330 (1918).

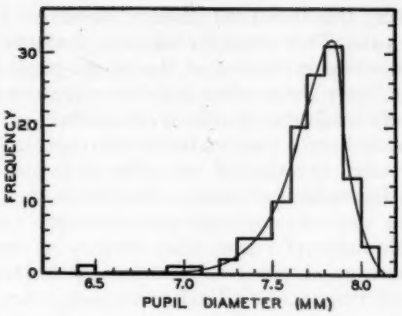

FIG. 2. The distribution of pupil diameter, as determined by 120 observations, in the absence of retinal illumination. The eye is fully dark adapted at the time of each observation.

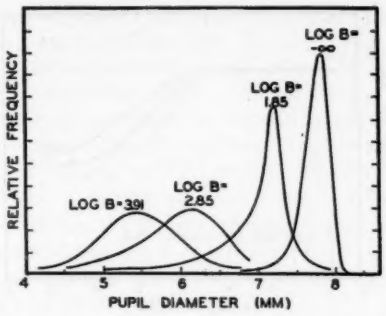

FIG. 3. Distribution of diameter for foveal stimulation. The curve for $\log B = 3.91$ is completely symmetrical.

Several thousand "still" pictures were made in this manner on 35-mm fine grain film (Fig. 1). These could be usefully enlarged until the dimensions of the eye on the photograph were twice those of the actual eye. The variations in magnification of the image of the pupil in the photographs due to the enlarging process and shift in the eye from central to peripheral vision were corrected for by measuring the fixed diameter of the sclero-corneal margin in the enlarged photographs.

RESULTS

The pupil diameter is not only affected by stimulation of the retina. There are mechanical factors, psychic influences, the ciliospinal reflex, and so forth. Probably one of the chief difficulties the author experienced was keeping the lids open without strain at the time the photoflash was to go off. Such a strain would tend to constrict the pupil (orbicular reflex). Because of the many factors which can affect pupil diameter, it is not too surprising to find that, for given stimulus, photographs taken randomly over a period of months exhibit a dispersion in pupil diameter.

Probably the most interesting case is shown in Fig. 2. The histogram and smoothed curve represent the distribution of pupil diameters taken from 120 photographs made in the absence of retinal illumination, each photograph being taken after thorough dark adaptation. As all of the photographs showed sharp edges for the border of the pupil, it seems reasonable to suppose that the spread in the distribution could not be due to the action of the flash bulbs. These observations were made in groups of five, the 120 photographs being distributed over a period of several years. There was no observable trend in pupil diameter during the years for this case of zero illumination so that the spread in the distribution cannot be attributed to a systematic change in the eye with age.

Figure 3 shows the distribution of pupil diameter for four brightnesses. The distributions are unsymmetrical when the mode is near the maximum possible diameter, but become increasingly symmetrical as the mode departs from this barrier with increasing brightness. These curves were smoothed from histograms.

In Fig. 4 is shown the variation of the mode of the pupil diameter with logarithm (base 10) of the brightness in mcm. Since the illumination is known to at least one percent and the logarithm of the brightness is the independent variable, the brightness must be considered as precisely known and all fluctuations in pupil diameter ascribed to the physiological behavior of the pupil. The circles represent the modes of the observations made at the indicated brightnesses, in the year 1943. The second set of data in Fig. 4 indicated by the crosses was the result of an attempt, in 1946, to determine the sharpness of the bend in the curve by interspersing more observations. What emerged, however, was the fact that the sensitivity of response of the author's photo-pupil reflex had decreased to a marked extent during the three-year interval.

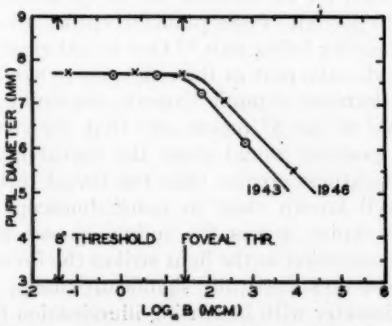

FIG. 4. Foveal vision. Pupil diameter (mode) plotted against the log of the brightness. The circles represent the modes of 160 observations made in 1943. The crosses represent 300 observations made in 1946. The thresholds for the author's eye are marked off in the abscissa.

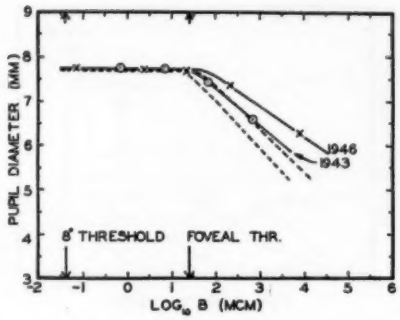

FIG. 5. Peripheral vision. The solid curves show the pupil diameter variation with log brightness. The corresponding foveal curves of Fig. 4 are shown as broken curves.

Figure 5 shows the corresponding observations at the 8° periphery (temporal region on the retina). These observations were scattered among the foveal observations so as to make comparison statistically valid. Again the three year shift in sensitivity is apparent. But in either case, the reflex shows considerably less sensitivity in the 8° periphery than in the fovea.

The important observations concerning the data of Figs. 4 and 5 are as follows. First, let us consider the case of the foveal stimulus. If, as has been suggested by some writers, neurons other than the cone receptors were responsible for initiating the reflex, one would expect that the reflex might function before the target could be seen (centrally). The evidence does not support this hypothesis, at least for the author's eye. When the light could not be seen, there was no significant change in the pupil diameter with change in illumination. When the brightness exceeded threshold, the reflex functioned. This suggests that the cones in the fovea were responsible for initiating the reflex, although probably a small fraction have the proper neural connections with the reflex pathway. Next, let us consider the 8° position. Here, about 10 percent of the photo-receptors are cones, the remainder being rods.[2,3] One would expect that if the rods take part in the reflex, there would be a steady decrease in pupil diameter beginning at the threshold of the 8° region and that the curve for the 8° position would cross the foveal curve at some brightness greater than the foveal threshold. It is well known that an ophthalmoscope light, when sweeping across the retina, causes marked pupil constriction as the light strikes the fovea. The data show no statistically significant change in the pupil diameter with increasing illumination for the 8° position until the brightness reached a value exceeding the threshold already observed for the foveal case. This strongly suggests that the cones dominate the initiation of the photo-pupil reflex. That is, the rods or other possible receptors play a relatively insignificant role in the reflex.

All this sounds reasonable in retrospect. For, if the primary function of the reflex is to assist the eye during sudden changes in brightness, and these changes are of no consequence except in the photopic range of vision, then there is no need for the photo-pupil reflex to function in the scotopic range of brightness. On the contrary, the reflex would lower visual efficiency if it operated in the scotopic range of vision.

SUGGESTIONS FOR FURTHER INVESTIGATION

The observations described above should be considered as only preliminary, for there is no guarantee that the author's eye is typical. Before generalizing the conclusions, the experiments should be repeated on a group of persons and should be extended to include monochromatic stimuli.

In regard to extending the brightness beyond $\log B = 4$ (B is brightness in mcm), it may be mentioned that observations made in this region of brightness suggested that the reflex for the foveal case became saturated with a pupil diameter of about 4.5 mm and $\log B \sim 6$. But the reflex began to function again beyond $\log B \sim 6.5$. A complete investigation of the phenomenon of saturation could yield information concerning the distribution of those receptors controlling photo-pupil reflex.

If we suppose that some of the cones and none of the rods initiate the photo-pupil reflex, it would be highly interesting to examine the behavior of the reflex in the total color blind, particularly those with a central scotoma and presumably possessing only rod vision.

Another approach to the study of the reflex is to compare the rate of change of pupil diameter with

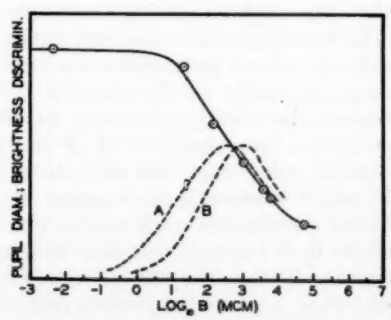

FIG. 6. Solid curve shows Reeves' data for pupil diameter corrected for pupil area and Stiles-Crawford effect. The broken curves, A and B, are two brightness discrimination sensitivity curves from Houstoun. (The ordinate units have been left arbitrary.)

[2] S. L. Polyak, *The Retina* (University of Chicago Press, Chicago, Illinois, 1941).

[3] G. A. Osterberg, Acta Ophthalm., Suppl. 6 (1935).

the brightness discrimination sensitivity. In this case, the principal variation from the experimental method just described would be to prepare the target so as to present to the eye a circular disk whose two halves are unequally illuminated, as in a photometer. The discrimination sensitivity of the eye is measured and at the same time the pupil photographed. If the cones are responsible for initiating the reflex, the maximum rate of change in diameter with $\log B$ should coincide with the peak brightness discrimination sensitivity. Separate data for brightness discrimination sensitivity and for pupil diameter variation with brightness are available in the literature. But what is needed is a simultaneous recording of the two sets of information on the same observer, the pupil being photographed just as the observer finishes a setting for brightness discrimination. To show that there is some promise to this approach, Fig. 6 is included showing on the same graph a plot of Reeves' pupil diameter data[1] corrected for area of pupil and the Stiles-Crawford effect[4] together with a plot of Houstoun's[5] data on brightness discrimination sensitivity.

In conclusion, the author wishes to acknowledge the cooperation of Dr. C. C. Schmidt, Head of the Physics Department, through whose efforts this research was made possible.

[4] W. S. Stiles and B. H. Crawford, Proc. Roy. Soc. (London) 112, 428 (1933).

[5] R. A. Houstoun, *Vision and Colour Vision* (Longmans, Green and Company, New York, 1932).

Erratum: The Course of Foveal Light Adaptation Measured by the Threshold Intensity Increment

[J. Opt. Soc. Am. 39, 172 (1949)]

HOWARD DEHAVEN BAKER
Department of Psychology, Columbia University, New York, New York

THROUGHOUT the article the pupil size was referred to as 0.856 mm². This should have been 2.70 mm². Figures accompanying the article giving brightnesses and photons calculated from the pupil area are correct; the error was made in writing.

Photographic Spectral Sensitometry with a Biprism

WILLOUGHBY M. CADY AND GARY GORDON
U. S. Naval Ordnance Test Station, Pasadena Annex, Pasadena, California

(Received January 24, 1949)

A method is described whereby spectral sensitometry is facilitated by the addition of a biprism to the optical train within a spectrograph. The modified instrument affords a spectrogram divided into three zones of relative intensities ½, 1, ½. In any spectral region, the characteristic curve of the emulsion can be determined from transmittance measurements on line images in all three zones. Although rather similar to the use of a step filter, the method has advantages over other techniques for homochromatic sensitometry.

IN determining photographically the relative intensities of a compact group of spectrum lines, it is necessary to calibrate the photographic material photometrically at the wave-length region in question. The problem of calibration is to ascertain, with a given exposure time, the relation between the relative intensity of the lines and the transmittance of the developed image. There are in use several procedures[1] in which nearby areas of the emulsion are exposed for the same time to various intensities in known ratios. The ratios are usually fixed by one of three alternative methods: (1) a sector disk rapidly rotating in the light path; (2) a step-weakener;[2] and (3) the use of lines known to have certain reliable intensity ratios.[3]

The methods enumerated above are imperfect. Method (1) is subject to error because the intermittent nature of the illumination only roughly imitates a reduction of the intensity; this error is most serious if the exposure time covers only a small number of rotations of the sector. Method (2), which rests on a preliminary determination of the spectral transmittance of the step-weakener, compounds the inaccuracies of this measurement with those of the photographic sensitometry. Method (3) is applicable only in spectral regions where reliable standards of relative line intensity exist. We will describe a step-weakener which is relatively free of these disadvantages.

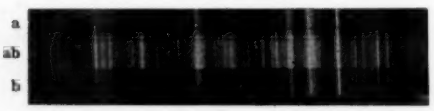

FIG. 1. Spectrogram taken with the biprism in the light path, showing the halved intensity of the outer zones.

Broadly speaking, the device is a pair of identical prisms having very acute refracting angles. These are installed in the spectrograph just before or just after the dispersing prism. They may be used with some grating instruments also. The prism faces are essentially normal to the rays; their refracting edges are horizontal, one being up and the other down. Effectively one-half the luminous flux traverses each prism. In our medium quartz spectrograph, the prisms are made of one piece of silica as a biprism with refracting angles 0°.13 and mounted just after the dispersing prism. The effect of the prisms is to deviate the rays in one-half of the aperture downward through a small angle and to deviate the remaining rays slightly upward. It results that each spectrum line is "telescoped," as in Fig. 1. Of the three zones so formed, the two marked *a* come from one of the prisms and the two marked *b* come from the other. The center zone is exposed to an intensity equal to the sum of the intensities falling on the outer zones. If the outer zones are equally bright, the center zone receives twice the intensity of either outer zone and sensitometry is possible. The intensity in the center zone is, except for reflection losses at the biprism, the same as with the unmodified instrument.

For accurate sensitometry, two conditions are necessary: first, each line must be free of gradual intensity changes along its length; and second, the outer zones must be equal. The first condition is met by the well-known precaution of focusing the source upon the collimating lens by a lens close before the slit. The second condition is met by empirically masking the halves of the biprism; this condition is rather sensitive to vertical displacement of either the biprism, the source, or the condensing lens.

The method of obtaining the curve relating the relative intensity I of the lines to the transmittance T of their developed images is not original, being based on a method[4] used in deducing the blackening curve from graded exposures of two neighboring lines whose intensity ratio is known and reliable.

[1] G. R. Harrison, J. Opt. Soc. Am. 19, 287 (1929).
[2] G. O. Langstroth and D. R. McRae, *Proceedings of Fifth Summer Conference on Spectroscopy* (John Wiley & Sons., Inc., New York, 1938) p. 105; Mary Banning, J. Opt. Soc. Am. 37, 686 (1947).
[3] G. H. Dieke and H. M. Crosswhite, J. Opt. Soc. Am. 33, 425 (1943).
[4] J. Raynor Churchill, Ind. Eng. Chem. A16, 653 (1944).

In the desired spectral region, one measures T in all three zones for each of several lines having a variety of intensities. Then on a graph labelled with coordinates T_1 (for the center zone) and $T_{\frac{1}{2}}$ (for the outer zones averaged) one indicates each spectrum line by a point and draws a smooth curve through these points as in Fig. 2. Intensity 32 may be arbitrarily assigned to the strongest line. In our example the transmittance $T_{\frac{1}{2}}$ of this line in the outer zones is 3.7 percent. To a hypothetical line in the center zone having this transmittance we may then assign intensity 16. The transmittance at the latter wave-length in the center zone is read from the graph and the step wise process may be iterated to intensities 8, 4, 2, etc. When this process is completed, the values of $\log I$ may be tabulated against $\log T_1$ as in Table I.

TABLE I. Values of T_1 and $T_{\frac{1}{2}}$ read from Fig. 2, giving the transmittances of lines of known relative intensities.

I	T_1	$T_{\frac{1}{2}}$
32	1.5	3.7
16	3.7	8.0
8	8.0	16.4
4	16.4	31.0
2	31.0	53.0
1	53.0	

In case the outer zones are not adjusted to precisely equal transmittances, it becomes important to know how seriously the maladjustment affects the accuracy of the method. The use of the arithmetic average of the transmittances in this case turns out, on analysis, to be very good, in the sense that $T_{\frac{1}{2}}$ so computed is almost exactly the transmittance that would correspond to a half-intensity exposure. The error incurred can be estimated for the straight part of the characteristic curve, which has slope $-\gamma$. It can be shown that for this part of the characteristic, $T_{\frac{1}{2}}$ should in principle be computed from the averaged values of $T_{\frac{1}{2}}^{-1/\gamma}$ of the outer zones, and that the error incurred in using the simple average is approximately $\delta^2 T_{\frac{1}{2}}(\gamma+1)/8\gamma$, where δ is the difference between the two outer transmittances divided by one of them. For $\gamma=1.5$, the factor $\delta^2(\gamma+1)/8\gamma$ is only 0.21 percent when δ is as great as 10 percent.

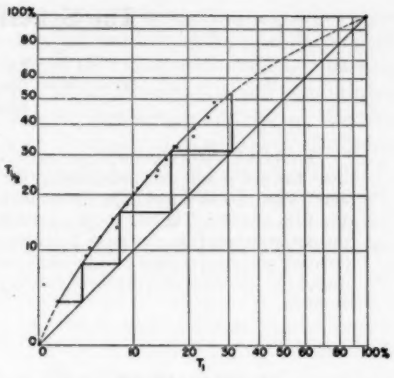

FIG. 2. Transmittances T_1 and $T_{\frac{1}{2}}$ of eighteen lines in the region 3950–4150A. The subscripts refer to exposures made at full intensity and half intensity.

It sometimes happens that the points on the $(T_1, T_{\frac{1}{2}})$ graph are so crowded toward large (or small) transmittance values that the steps cannot be accurately drawn in. In such a case it has been found convenient to use identical non-uniform scales for T_1 and $T_{\frac{1}{2}}$, chosen to bring the steps into a region where they are larger and easier to draw and read accurately. Such scales are conveniently provided by log-log paper, if its graduations 1, 2–9, 10, 11 are relabeled for example 0, 10–80, 90, 100 percent (useful where the spectrum is heavily exposed, see Fig. 2) or 100, 90–20, 10, 0 percent (useful where the spectrum is faint). In any case the points labeled (0, 0) and (100, 100) should be marked on the graph since the curve must pass through them.

It is believed that the method of telescoped spectrum lines described above may be widely useful. It is fairly rapid; it avoids reciprocity and intermittency failure. It is self-checking, in that failure of uniform illumination of the slit is obvious from the inequality of the outer zones; and the spectrograph may be restored to a conventional status by simply raising the biprism out of the optical path. The method is in use in research on the time-variations of the spectra of single sparks and exploding wires.

The Spherical Aberration of the Eye

M. KOOMEN, R. TOUSEY AND R. SCOLNIK
U. S. Naval Research Laboratory, Washington, D. C.

(Received January 27, 1949)

The spherical aberration of the eye was measured by placing a series of centered annular apertures over the eye pupil, and determining the optimum spectacle correction for each aperture. A "double star" was used as a test object. Accommodation was controlled by reflecting a second test object into the field of view. The three eyes examined had positive (undercorrected) spherical aberration when unaccommodated; in one case 2 diopters at the pupil margin. The aberration was reduced with increasing accommodation and in one case became negative at high accommodation. Homatropine reduced the spherical aberration of two of the three eyes examined. A historical review of previous work is given.

INTRODUCTION

IN the course of experimental work on night vision and night myopia a search of the literature was made for data on the spherical aberration of the eye. From the published results it was evident that the aberration varied considerably among individuals, and that it would have to be measured for the individuals whose night vision was to be tested. Measurements described in the literature were usually made on selected quadrants of the pupil, or point by point along one or two selected meridians, and therefore gave a somewhat incomplete picture of the aberration of the entire eye. Moreover, a cycloplegic was used in many of these measurements, and there appeared to be a possibility that this changed the spherical aberration. Therefore, an experiment was devised to avoid these difficulties. A better over-all picture of the aberration was gotten by making the measurements on centered annular zones of the pupil, rather than points or quadrants, and the use of cycloplegia was avoided by controlling the accommodation by optical means. The results of the experiments with three subjects showed that positive (undercorrected) spherical aberration is present in the normal unaccommodated eye. The aberration was reduced when the eye accommodated and, in two individuals, became zero or even negative (overcorrected) at high values of accommodation. The use of homatropine changed the spherical aberration in two cases. The history of the subject will be covered briefly since it does not appear to have been assembled previously.

I. HISTORICAL

The presence of spherical aberration in the human eye appears to have been recognized at least as early as 1801. In that year Thomas Young[1] demonstrated the aberration by means of a simple experiment. He extended a thread from a point close to one eye, nearly along its visual axis, to a considerable distance away. Through four narrow slits placed directly over his eye pupil, he observed four separate images of the threads. All four images appeared to intersect at a single point, presumably his far point, when his accommodation was relaxed. When he accommodated for near vision, however, the two images formed through the outer slits appeared to intersect beyond the point where the inner ones crossed. Apparently Young's eye had little spherical aberration in its relaxed state, but exhibited negative or overcorrected aberration when accommodated.

In a similar experiment Volkmann,[2] in 1846, looked through four small openings placed across the eye pupil at a pin which he moved to different distances. As a measure of the spherical aberration, he determined the distance at which the pin was in sharp focus, using first the two more centrally located openings, and then the two located more laterally. Of nine observers tested, he found five with positive spherical aberration and four with negative aberration.

In his original papers Helmholtz,[3] in 1866, wrote, "the experiments of Young and Volkmann would undoubtedly have indicated the nature and magnitude of the spherical aberration of the eye;" but, "in most meridians of the ordinary eye the points of intersection of the refracted with the central ray do not form a continuous series at all, so that here the conception of spherical aberration does not apply." In his supplement, however, Helmholtz included the results of investigations by Donders (1861) who "moved about in front of the eye a small perforated screen so as to let light go sometimes through one sector of the lens and sometimes through another···each sector of the lens converged the rays approximately to a focus, but the focus was not the same for different sectors. And, moreover, the focusing was not quite accurate

[1] T. Young, Phil. Trans. 91, 69 (1801).

[2] A. W. Volkmann, Wagner's Handworterbuch der Physiologie (Vieweg and Son, Brunswick, 1846), p. 293.

[3] H. von Helmholtz, *Physiological Optics*, 1909 edition; translated by J. P. C. Southall (Opt. Soc. Am., 1924), pp. 199–201.

even for a single sector, the rays near the axis having apparently a longer focal length than those traversing the peripheral parts of the lens."

The characteristic skiascopic appearances produced by symmetrical positive and negative aberrations were first described by Jackson[4] in 1885. Within the next three years he measured skiascopically the aberration in one hundred eyes.[5] Of these, seventy-eight had one-half diopter or more of positive aberration, thirteen were within one-quarter diopter of zero, and only nine had negative aberration. The negative aberrations were largely associated with conical corneas, or with lens conditions which often led to nuclear cataract.

In 1898 Tscherning[6] devised his "aberrascope" for an investigation of the effect of accommodation upon spherical aberration. He ruled opaque lines on the flat side of a plano-convex lens. With the lens in front of his eye, he observed a distant luminous point. The opaque lines cast shadows in the blur circle produced on the retina by the "artificial myopia." From the direction of the curvature of the shadows Tscherning concluded that the aberration was positive when his eye was relaxed, tending toward the negative as his eye accommodated.

Gullstrand,[7] in 1909, described as mathematically impossible Tscherning's interpretation of the phenomena he had observed with his aberrascope, and declared that "subjective stigmatoscopy, employed as a scientific method, demonstrates a positive aberration inside the optical zone even with the most powerful accommodation." Nevertheless, Gullstrand acknowledged that "skiascopic investigation shows that in many instances a change of sign occurs in the peripheral aberration during accommodation."

In 1921 Ames and Proctor[8] measured the spherical aberration along two meridians in the relaxed eye. They directed at the eye two beams of light of small cross section. One beam passed through the center of the pupil while the second was made to pass consecutively through a series of points along a given meridian. The two beams were made to intersect on the retina by an adjustment of the inclination of the second beam. The spherical aberration was calculated from the angle between the beams. In three eyes measured, Ames and Proctor found that from the center of the pupil to a radius of about 1.5 mm the eye became increasingly myopic; that is, it showed positive spherical aberration. From 1.5 mm to the pupil margin the aberration decreased, tending to become zero again at the very edge. The maximum observed values of the aberration were between 0.4 and 0.8 diopter.

In 1925 Pi[9] made further use of the skiascopic method in a study of the spherical aberration of fifty patients at the Poliklinik in Vienna. Dividing the pupil area of the homatropinized eye into a central zone and four peripheral quadrants, he measured the difference in refraction between the central zone and each quadrant in turn. Pi found that all except one of his subjects possessed positive aberration, the individual magnitudes ranging from 0.25 to 5.0 diopters depending upon the subject and upon the quadrant. Most frequently, however, the peripheral quadrants were found to be only 1.0 to 1.5 diopters myopic relative to the center. In not a single instance was the aberration the same all around the periphery. Pi observed that the sequence of the magnitudes of the refractive errors among the four quadrants (superior, 1.89 D.; nasal, 1.50 D.; temporal, 1.37 D.; and inferior, 1.04 D.) was directly opposite to the sequence to be expected from a consideration of the measurements of corneal curvature made by Aubert, Sulzer, and Erikson, and by Gullstrand.

In 1930 Stine[10] made a similar investigation of the refraction by quadrants in 277 non-pathological eyes. He found the same sequence among the quadrants that Pi did with respect to the amount of the refractive error. Stine, however, measured differences between quadrants as high as 8 diopters, the greatest difference between periphery and center being 7.5 diopters. Some degree of aberration was found in every eye; 66 percent had some positive aberration, 46 percent had some negative aberration, and 17.5 percent showed mixed (scissors) aberration. Most of the negative and mixed aberration was found in young subjects, with not one instance of entirely negative aberration in an individual over nineteen years of age. Stine attempted to reconcile his data, and those of Pi, with Gullstrand's measurements of corneal flattening by placing upon the lens rather than upon the cornea the major responsibility for the resultant positive aberration of the entire eye.

More recently, in 1941, Otero and Duran[11] investigated the aberrations of the eye in the course of their studies of nocturnal myopia. Looking through each of a series of circular artificial pupils of different diameters, they determined the spectacle correction which yielded the best visual acuity while observing a grating type target. Their results

[4] E. Jackson, American J. Med. Sciences **89**, 404 (1885).
[5] E. Jackson, Trans. Am. Ophth. Soc. **5**, 141 (1888).
[6] M. Tscherning, Optique Physiologique (Carre et Naud, Paris, 1898), pp. 98 and 158.
[7] A. Gullstrand, *Supplement to Helmholtz's Physiological Optics*, 1909 edition; translated by J. P. C. Southall (Opt. Soc. Am., 1924), p. 437.
[8] A. Ames and C. A. Proctor, J. Opt. Soc. Am. **5**, 22 (1921).
[9] H. T. Pi, Trans. Ophth. Soc. of United Kingdom XLV, 393 (1925).
[10] G. H. Stine, Am. J. Ophth. **13**, Ser. 3, 101 (1930).
[11] J. M. Otero and A. Duran, Anales de Física y Química **38**, 236 (1942).

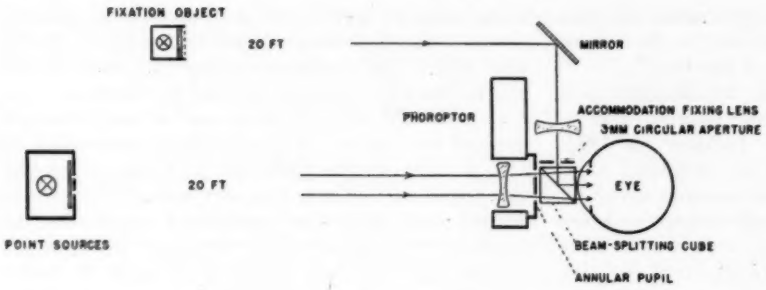

FIG. 1. Apparatus for measurement of the spherical aberration of the eye.

indicated the presence of spherical aberration somewhat similar to that found by Ames and Proctor. However, since the spectacle corrections were determined not zone by zone, but for circular areas of progressively increasing size, the results do not give directly the spherical aberration of the eye. They yield, instead, a value of aberration which is less than the maximum zonal aberration and which is determined by the way the various zones of the eye pupil contribute to the retinal image.

In 1944 Byram[12] measured the spherical aberration of his eye as a part of the study of the light distribution in the image on the retina. He reported that his eye became increasingly myopic from the center to the margin of the pupil, the aberration at the margin amounting to 1.3 diopters. Byram did not, however, describe his experimental method.

Von Bahr[13] in 1945 gave the results of his measurements of the spherical aberration of 30 eyes. He measured the aberration along the horizontal and vertical meridians of the pupil, using pairs of small pinholes to sample various zones along a meridian. He found that most of the eyes exhibited positive spherical aberration, the zones of the pupil becoming gradually more myopic from the center of the pupil outward. In a few eyes negative aberration was present. Also, a small number of eyes exhibited negative aberration in one-half of a meridian only. In such cases the aberration was usually small. Of all the eyes measured, there was none in which the aberration was completely symmetrical. In the most extreme cases of positive aberration, the pupil margin was measured to be about 3 diopters myopic relative to the center. However, the mean value for all eyes and all observations appeared to be about 0.5 diopter.

Ivanoff[14,15] in 1947, modified the Ames and Proctor technique to permit the use of vernier acuity in making settings and made measurements out to 2.5 mm from the center of the pupil. He found that along a single meridian the outer zones of the average eye exhibited approximately 0.9 diopter of positive spherical aberration when in the relaxed state, and 1.25 diopters of negative spherical aberration when accommodated 3.0 diopters. When accommodated at 1.5 diopters, there was practically no aberration present. While the magnitude of the aberration varied from individual to individual, the amount of the accommodation necessary to eliminate the aberration was about the same for all individuals. As plotted by Ivanoff, the curves show the major change in refractive power occurring within 0.5 mm of the center of the pupil, with little additional change beyond this zone. This is unexpected in a reasonably simple image-forming system, and it does not agree with our findings or those of other investigators.

This account of the developments in the measurement of spherical aberration shows the general agreement among the investigators that the average eye in the relaxed state is characterized by positive aberration. Some investigators found, in addition, that with accommodation the aberration is diminished, and sometimes it even becomes negative. It is with respect to the magnitude of the aberration and to its change as one considers more and more marginal pupil areas that there is less agreement among investigators.

Much of the disagreement can undoubtedly be attributed to real differences in the characteristics of the eyes examined. Measurements made by individual experimenters using a single technique with many subjects would bear this out. In addition, however, there were considerable differences in the experimental methods employed. Several investigators divided the eye pupil into quadrants. Others measured very small areas along a few selected meridians. Otero and Duran's data referred to circular pupils of different areas. As a result nearly every investigator measured a slightly different aspect of the spherical aberration of the eye, and the results varied accordingly.

II. EXPERIMENTAL

The experiments to be described were made in an attempt to measure spherical aberration in such

[12] G. M. Byram, J. Opt. Soc. Am. **34**, 590 (1944).
[13] G. Von Bahr, Acta Ophth. **23**, 1 (1945).
[14] A. Ivanoff, Rev. d'Optique **26**, 145 (1947).
[15] A. Ivanoff, J. Opt. Soc. Am. **37**, 730 (1947).

a manner as to relate it most directly to vision, in order to determine its true role among the causes of nocturnal myopia. For the purpose of measurement, the eye pupil was subdivided into a central circular area of two millimeters diameter and five to seven distinct annular areas of increasing mean diameter extending to the limit imposed by the fully dilated pupil. Measurements consisted, essentially, of determining the spectacle lens power which produced most acute vision while the observer was using, in turn, each of the chosen pupil areas. The spectacle corrections for the annular zones, considered relative to the correction required for the small central area, gave the magnitude and the sign of the aberration. The measurements were made under conditions of controlled accommodation ranging from complete relaxation to the maximum the observer could exert without undue distress.

The arrangement of the apparatus is shown in Fig. 1. The observer looked through the right eye-aperture of an additive phoroptor. The phoroptor is an instrument which permits spectacle lenses of powers varying by quarter-diopter steps to be placed before the eye by turning a lens-bearing disk. Supplementary lenses provide eighth-diopter steps when desired. The annular apertures used for dividing the eye pupil into the selected zones were placed in turn in front of the right eye-aperture of the phoroptor. The annular apertures consisted of transparent rings in an opaque background printed on thin photographic film. The dimensions of the annular apertures are shown in Fig. 2, with the mean diameters noted above. The mean diameter is defined as that diameter which divides the transparent zone into equal inner and outer areas. The target observed through the annular aperture and phoroptor consisted of a distant pair of point sources whose separation was held just above the resolution threshold for each annular zone. The intensity of the double star source was high enough for cone vision.

To fix the state of accommodation of the eye, a second target was introduced by means of a small beam-splitting cube placed between the phoroptor and the observer's eye. This accommodation-fixing target consisted of an opaque sheet of photographic film with four transparent letters printed upon it. The target was illuminated from the rear through a sheet of ground glass. The observer thus saw the double star source surrounded by four bright letters. A spectacle lens of the proper power was inserted into the optical path of the accommodation-fixing target in order to place the target at any desired optical distance from the eye. The accommodation was therefore stimulated for this distance. The maximum accommodation which the observer could attain with this arrangement was somewhat less than could be reached with binocular vision.

The accommodation-fixing target was viewed through an artificial circular pupil of 3 mm diameter placed on the side of the cube in the path to the accommodation-fixing target. This circular pupil was used in order to define the portion of the eye pupil which determined the accommodation. The 3 mm diameter was found to be the most satisfactory size because it produced a sharp retinal image and avoided excessive depth of focus.

The eye pupil was centered relative to both the annular ring and the 3 mm pupil by the observer himself. A negative lens was set up in the phoroptor so that the double star source was thrown out of focus and looked like a large single bright ring. The observer then positioned his head and eye so that the ring appeared as circular and as uniformly bright as possible. This position was maintained by the use of a bite plate. The adjustment was checked during a set of determinations by throwing the double star out of focus and noting the symmetry of the blur image.

A check on the circular symmetry of each eye was obtained by changing toward more positive or negative values the power of the lens used to throw the image of the points out of focus. As the powers changed, the cone of rays entered the eye through different zones of the eye pupil because of the separation between the annular pupil and the eye. Once the pupil was centered by the observer, a nearly circular out-of-focus blur circle was obtained for all powers, indicating that the aberrations of the eye were approximately independent of the meridian.

The Stiles-Crawford effect was sometimes of aid in the centering. Decentering of the 2.1 mm central pupil caused the brightness and the sharpness of the image, whether in-focus or out-of-focus, to be reduced. Decentering of the annular apertures

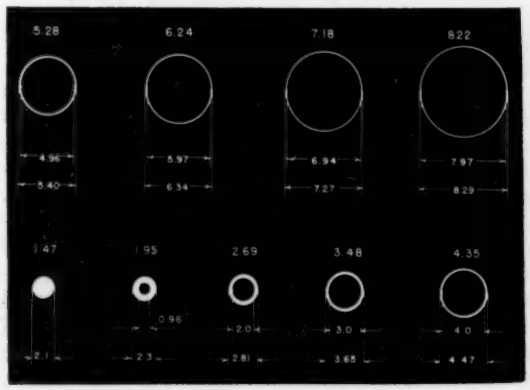

FIG. 2. Annular pupils for measurement of the spherical aberration of the eye. Dimensions are in mm. The mean diameter is given above each annulus.

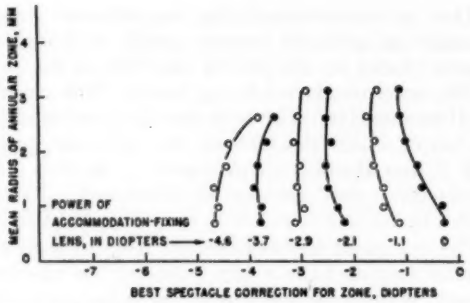

FIG. 3. Spherical aberration curves of the right eye of subject M.K. (emmetrope).

caused in-focus images to increase in brightness as one edge of the annulus approached the pupil center. With out-of-focus images, the side of the blur ring produced by the side of the annulus nearer the pupil center increased in brightness while the brightness of the opposite side decreased.

In making a measurement, the observer held the fixation letters in sharp focus while manipulating the phoroptor lenses until the best resolution of the double star was obtained. Starting with the smallest pupil, the observer repeated this procedure with larger and larger annular apertures, up to the limit set by the natural pupil. Because the annular apertures were constructed of equal areas, the intensity of the double star was increased to compensate roughly for the Stiles-Crawford effect as larger annular rings were used.

The writers themselves served as observers. Each observer had 20/20 vision; M.K. and R.S. without correction, R.T. with his −4.25 diopters prescribed correction. Their ages were respectively 29, 30 and 39. No other observers were used because it was believed that untrained observers would require a considerable amount of instruction and practice before they could produce consistent results.

III. RESULTS

The results of the experimentation are shown in the spherical aberration curves for various states of accommodation plotted in Figs. 3–5. The ordinates are the mean radii of the annular zones. The abscissae are the best spectable corrections for the zones. Below each curve is indicated the power of the "accommodation-fixing" lens; that is, the power of the lens inserted into the path to the accommodation-fixing target. The powers indicated are referred to the cornea, and indicate the refractive effect, in diopters, of the lens measured at the vertex of the cornea. The best spectacle corrections for the zones, plotted as abscissae of the curves, are also for lenses placed at the cornea. Positive spherical aberration is indicated by points to the left of the point for the central zone, and negative aberration by points to the right. The plotted points are the average of several independent determinations taken over a period of nearly a year. The data on different runs agreed to within $\pm\frac{1}{4}$ diopter except for an occasional erratic setting. In Figs. 3 and 5, the curves marked zero represent the spherical aberration of the relaxed eyes of M.K. and R.S., respectively. The curves for myope R.T. are shown in Fig. 4, and the curve marked −4.6 represents the spherical aberration of his relaxed eye. The curves to the left of the relaxed accommodation curves show the spherical aberration for various states of accommodation. If, with accommodation, the spherical aberration remained unchanged, or, to put it another way, if all zones of the eye accommodated equally, the curves would simply be displaced to the left by amounts equal to the accommodation and would suffer no change in shape. Generally this did not occur, and the curves illustrate the relative change in refractive power of the several zones of the eye as the eye accommodated.

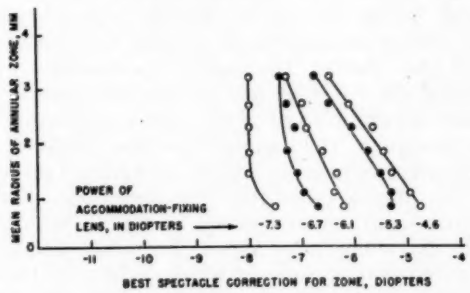

FIG. 4. Spherical aberration curves of the right eye of subject R.T. (myope, −4.25 *D* prescribed correction).

Since the accommodation-fixing target was viewed through an aperture of 3 mm diameter (mean radius of 1 mm), it may be expected that the curves would have an abscissa at the 1 mm mean radius annular zone equal to the value of the accommodation-fixing lens. This does occur to within 0.2 diopter and is a measure of the success with which the desired accommodation was achieved by the experimental arrangement. However, it will be noted that in all 3 figures the curve for relaxed accommodation lies too close to the other curves. This indicates that all three observers had some uncorrected myopia. M.K. and R.S., for example, were not quite emmetropic, but were myopic by about $\frac{1}{4}$ diopter. Figure 5 shows that the myopia of R.T. was not fully corrected by his −4.25 diopter prescribed correction, and that his full correction was nearly five diopters.

Observer M.K. showed a maximum of about 0.75 diopter of positive spherical aberration when

his eye was unaccommodated (Fig. 3), the marginal zone being myopic by that amount relative to the center. The aberration decreased with increasing accommodation and became nearly zero for about three diopters accommodation. With still greater accommodation the marginal zones showed definite negative aberration. The outermost zone could not be measured for more than 3 diopters accommodation because of the accompanying pupil constriction.

Observer R.T. showed positive aberration with no accommodation, the maximum being 1.6 diopters for the marginal zone, (Fig. 4). This was reduced to about 0.5 diopter with an accommodation of 2 diopters, the maximum he could exert consistently. It can be seen that the amplitude of accommodation of his marginal zones is considerably less than for the inner zones of his eye.

Observer R.S. (Fig. 5) had a maximum of 2 diopters of positive aberration in the marginal zone with relaxed accommodation. This aberration was reduced considerably, although by no means eliminated, when he accommodated strongly. Indeed, the shapes of the curves out to a 2-mm radius remained substantially the same for all except the strongest state of accommodation. At the strongest state of accommodation the aberration was reduced to 0.7 diopter in the 3-mm radius zone. The aberration of more marginal zones could not be measured, because of the pupil constriction accompanying the accommodation.

Because of the considerable difference in the magnitude of the spherical aberration found among our three observers it was not considered meaningful to make a close comparison with the results reported by previous investigators; qualitatively, however, the agreement was satisfactory. Positive aberration was found for the relaxed eye; the aberration was reduced as the eye accommodated; and, in one instance, negative aberration was reached at high accommodation.

Within the central zone of 2.1-mm diameter it was not possible to make measurements by means of the annular zone method because the depth of

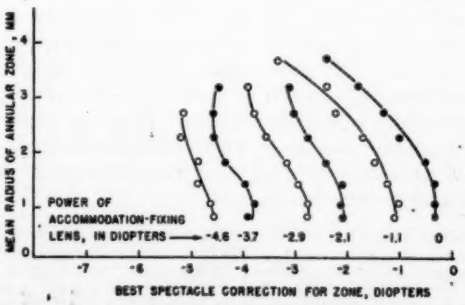
MEAN RADIUS OF ANNULAR ZONE, MM
POWER OF ACCOMMODATION-FIXING LENS, IN DIOPTERS
-4.6 -3.7 -2.9 -2.1 -1.1 0
BEST SPECTACLE CORRECTION FOR ZONE, DIOPTERS

FIG. 5. Spherical aberration curves of the right eye of subject R.S. (emmetrope).

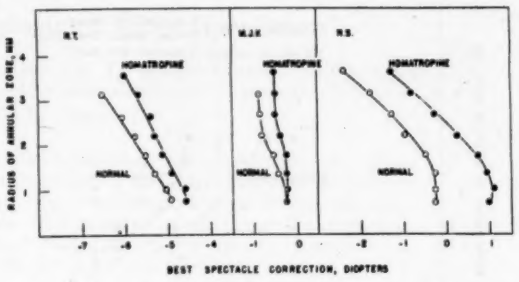

FIG. 6. The spherical aberration of the homotropinized eye compared to that of the normal unaccommodated eye.

focus of the eye was too great and because the acuity was reduced by diffraction. There was, however, no reason to expect appreciable aberration in this area for a simple lens system like the eye. Ivanoff,[14] nevertheless, reported large aberrations across the center of the pupil, measured by introducing a pair of pencils of light into the eye and arranging them so that the vernier acuity of the eye was utilized in making settings. The pencils were of 0.5-mm cross section at the pupil. The average aberration for the unaccommodated eye was found by him to be 0.75 diopter positive between the center and a point 0.75 mm out. This aberration was reduced to zero for 1.3 diopters accommodation and became 1.4 diopters negative for 3 diopters accommodation.

We have repeated Ivanoff's experiment in modified form using pencils of 0.2-mm cross section at the eye. The results we obtained across the horizontal meridian were in good agreement with our data obtained with annular pupils. No aberration was detected, however, inside the 1.5-mm diameter zone, with or without accommodation. The lower limit which could be detected was about 0.2 diopters for entering pencils separated by 0.75 mm. Our results therefore do not agree with those of Ivanoff for the center of the eye pupil.

It has been noted in the historical review that many determinations of spherical aberration have been made while the eyes were under cycloplegia. This was done to dilate the pupil and to relax the accommodation, and it was assumed that the aberration under these conditions remained unchanged. In order to determine whether cylcoplegia might change the spherical aberration, our measurements were repeated for the same observers after instillation of two drops of 2 percent homatropine into each eye. The resulting aberration curves are plotted in Fig. 6, along with the curves for relaxed accommodation, obtained before the homatropine was instilled. A comparison of the curves in each pair shows that in the case of M.K. and R.T., the spherical aberration was somewhat reduced by homatropine, the curves becoming more

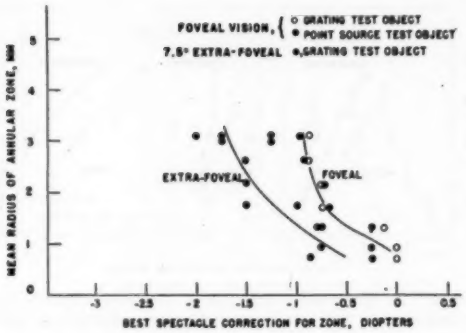
FOVEAL VISION, ○ GRATING TEST OBJECT
● POINT SOURCE TEST OBJECT
7.5° EXTRA-FOVEAL ●GRATING TEST OBJECT
MEAN RADIUS OF ANNULAR ZONE, MM
EXTRA-FOVEAL
FOVEAL
BEST SPECTACLE CORRECTION FOR ZONE, DIOPTERS

FIG. 7. The spherical aberration of the eye 7½° nasal from the fovea, compared with the aberration at the fovea.

nearly vertical. For the marginal zones, this reduction amounted to about 0.5 diopter. Also, the curve of R.T. was displaced slightly in the direction of hyperopia. For observer R.S., the homatropine did not change the shape of his spherical aberration curve measurably, but shifted the entire curve in the direction of hyperopia by approximately one diopter. Otero and Duran have also reported changes in the spherical aberration of their observers caused by the use of atropine, and they found one case where the drug made an observer hyperopic. On the other hand, the curves of Ames and Proctor seem to show no change in the aberration with the use of homatropine. Apparently cycloplegia has unpredictable effects upon the characteristics of the eyes of different individuals and does not necessarily leave the eye in its normal state of refraction.

An attempt was made to measure the extra-axial spherical aberration of the eye. Although such a measurement is complicated by the presence of other extra-axial aberrations, it was belived that the result would be of some significance, since it would be more directly related to night vision, which involves the extra-axial portions of the retina. The same experimental setup was used, but with a fixation target placed at a given angle from the test object. The measurements could not be made with accuracy because the extra-foveal acuity is low. Furthermore, it was difficult to obtain proper centering of the annular pupils for extra-foveal vision. A typical result of the measurements, using relaxed accommodation, is shown in Fig. 7. The retinal image was 7.5 degrees to the nasal side of the fovea. A grating type of test object was used. Each point represents a single measurement. In spite of the scatter, the extra-axial points indicate a possibly greater aberration than the axial aberration, a curve of which is reproduced for comparison. There is also an indication that the whole eye became 0.5 diopter myopic for vision at 7.5 degrees from the fovea.

In conclusion, it is apparent that when considering the effect of spherical aberration on vision, the aberration must be measured for the individual under examination. Furthermore, the experimental method should be as directly related as possible to the phenomenon of vision being investigated. For the study of night myopia, for example, the determination of the zonal aberrations of the eye by means of centered annular zones as here described appears to be a very suitable method. A close relation between night myopia and spherical aberration when measured in this fashion was found for our observers and will be discussed in another paper.

A New Spectrophotometer Employing a Glass Fery Prism*

W. C. MILLER, G. HARE, D. C. STRAIN, K. P. GEORGE, M. E. STICKNEY, AND A. O. BECKMAN
Research and Development Department, National Technical Laboratories, South Pasadena, California
(Received January 31, 1949)

The Fery prism, first described in 1910, has been used in relatively few applications, all of these apparently being in spectrographs. The present paper describes the application of a borosilicate crown glass Fery prism as the dispersive element in a monochromator covering the spectral range from 320 mμ to 1000 mμ. Performance characteristics of the prism are given. A new Beckman spectrophotometer employing the Fery prism is described along with its performance characteristics. A novel wave-length control mechanism and a new slit device are described.

INTRODUCTION

THE past ten years have seen the advent of a great many colorimeters and spectrophotometers incorporating a wide variety of principles both in the monochromators and the photometers. In view of the great number of designs which have been reported in the literature, it is surprising to find that a unique combination of conventional elements, which results in improved performance, has escaped attention until the present time.

This conspicuous oversight is in the choice of a dispersive element employed in monochromators. It is the purpose of this paper to describe this element and a new spectrophotometer in which it is incorporated.

DISPERSIVE ELEMENT

Although this is the first instance found in a careful search of the literature of the use of this type of dispersive element in a monochromator, the element itself is not new, but has been previously employed only in spectrographs. In 1910 Charles Fery[1] described a prism having curved rather than the conventional flat faces. The purpose of these curved surfaces was to endow the prism with the dual function of imaging and dispersion. This prism receives light diverging from the entrance slit and returns it, in the Littrow manner, re-imaged in a spectrum in the vicinity of the entrance slit.

The trail that led to the adoption of the Fery prism was a fairly direct one. Since 1941, the Beckman quartz spectrophotometer[2] has been employed in a wide variety of applications. Its wide wave-length range from 220mμ to 1000mμ, measurement precision, and other features made it a very versatile instrument. The high standards maintained made its cost comparatively high for certain applications. Thus, there has been an increasing demand for a medium-priced spectrophotometer with a more limited wave-length range, but possessing good photometric accuracy.

Glass or gelatin filters have been employed in many instruments as a means of limiting the radiation presented to the sample. The disadvantages of such instruments are three: They permit work at only a few wave-length regions; they inevitably have wide spectral band widths; and their optical transmission efficiency is low, and decreases with improved spectral isolation. They therefore offered little attraction in the design of the new instrument. Narrower pass bands are available with modern interference filters; however, to avoid stray light, they must be combined with glass filters.

Original ruled gratings, particularly of the echelette type, offer several advantages. A large amount of the incident light can be made to fall on the exit slit. The linear dispersion permits constant band width, and the problem of wave-length calibration is simple. However, there are two important objections to ruled gratings: The cost is prohibitive, and the stray light due to scattering and interference from other orders is so high that requisite purity of the exit beam can only be maintained throughout the spectrum with the aid of a battery of filters. Replica gratings, while less expensive than originals, have all of the stray light properties of originals plus an ample quota of their own.

Consideration was therefore directed toward the more promising prism systems. The conventional collimator and prism combinations were reviewed in considerable detail, but always the problem of the cost of two optical elements prompted one to seek a still simpler system. The only solution seemed to be a prism of the Littrow type, which could act as its own collimating and imaging element. The obvious answer was the much neglected Fery prism. A few simple computations and tests quickly indicated that here indeed was a simple element possessing all of the properties which were desired if means could be found to circumvent its astigmatism.

* Presented at the meeting of the Optical Society of America, Detroit, Michigan, October, 1948.

[1] Charles Fery, "A prism with curved faces for spectrograph or spectroscope," J. de Phys. **9**, 762 (1910).

[2] H. H. Cary and Arnold O. Beckman, "A quartz photoelectric spectrophotometer," J. Opt. Soc. Am. **31**, 682–689 (1941).

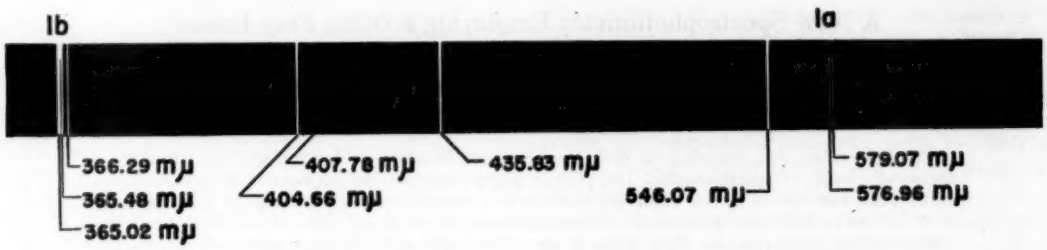

FIG. 1. Mercury spectrum photographed with a Fery prism. A portion of the mercury spectrum photographed by means of an F/18 Fery prism of 1-inch aperture, made of borosilicate crown glass. The doublet at *a* and the triplet at *b* give indications of the definition obtainable with a Fery prism when the entrance slit is of the proper curvature.

A detailed discussion of the principles involved in the Fery prism are reserved for a subsequent paper. Suffice it to say here that, by employing the prism in a way which avoids astigmatism, its chief optical fault, it is found both computationally and experimentally that the Fery prism is capable of competing successfully with the more complicated conventional systems. Its resolution is excellent, as is shown by the fact that an F/18 Fery prism of BSC-2 glass can clearly resolve the mercury doublet at 577 and 579mμ as shown at *a* in Fig. 1. The mercury triplet at 365.0, 365.5, and 366.3mμ are also clearly resolved as seen at *b* in Fig. 1.

Computations by standard ray tracing methods in the meridian plane of an F/18 prism show that the tangential coma is negligibly small, being only -7.4×10^{-6} at 300mμ, -5.2×10^{-5} at 589mμ, and -4.7×10^{-4} at 1000mμ. This is to be compared with the coma found in a comparable spherical mirror working at 10° between the incident and reflected beams, in which case the tangential coma amounts to $+4.9\times10^{-3}$ throughout the spectrum.

The only aberration with which the Fery prism is seriously afflicted is astigmatism. With a prism of 18-inch focal length, the astigmatism is such that a pinhole source is imaged as a line about three-quarters of an inch long. It is therefore essential that the images of all points of the entrance slit fall on the same straight line if maximum definition is to be obtained. Fortunately it is possible to employ an entrance slit of such curvature that the images of all its points produce this single straight exit image. As with all prisms, the required curvature is a function of wave-length, and an intelligent choice of the wave-length of curvature-match must be made from considerations of normal operation slit widths. A perfectly straight image can be produced at only one wave-length, but departures are slight over a large part of the wave-length range.

The image quality is surprisingly good when it is considered that the spherical first surface of the prism is only an approximation to the logarithmic spiral which theory decrees. Since spherical surfaces are entirely satisfactory for these elements they lend themselves admirably to quantity production methods.

DESIGN OF PRISM AND CARRIAGE

Consisting as it does of curved refracting and reflecting surfaces, the focal length of the Fery prism inevitably varies with wave-length, and mechanical means for shifting focal length with angle must be provided. If the position of the slits is fixed, as is essential for simple control of their width, the prism must be subject to translational and rotational motion simultaneously. The pivot point about which it would ideally rotate lies equidistant from the prism and the exit slit, and in the case of a prism of 18-inch focal length, this point lies about sixteen inches away. In the interests of compact design, a linkage of this length is out of the question.

The problem of a wave-length drive resolved itself into devising some compact mechanical system which would move the prism in the desired manner or some reasonable facsimile thereof. Several systems were studied. The most promising was a simple quadrilateral hinged at the four corners.

To understand the operation of this system, consider Fig. 2a showing a hinged parallelogram with one side, *A*, rigidly fixed. On the movable side, *B*, is mounted the Fery prism, *F*. As member *B* moves toward or from the slits, it is constrained by the parallelogram to move so that there is no rotation of the prism. There is a slight lateral displacement, but this will be ignored for the moment. If we now make the fixed side of the parallelogram shorter or longer than the moving side on which the prism rides, as in Fig. 2b, a certain amount of rotation will be imparted to the prism as the plate *B* carries it toward or away from the slit. The rate of rotation with respect to the rate of focal travel, is a function of the difference in length of the two sides of the quadrilateral. It is possible to find a relationship which causes the prism to rotate at very nearly the ideal rate. The slight lateral displacement of the prism mentioned above intro-

duces a slight error in focus, but this can be reduced to a minimum by careful adjustment and is largely absorbed in the depth of focus of the prism.

In practice the quadrilateral is a three-dimensional affair as shown in Fig. 3. Four, long beryllium-copper reeds are employed, one at each corner of the system. These reeds permit the required movement of the prism carriage and provide an inexpensive linkage between the moving members, allowing motion in one plane only.

The characteristics of all optical materials are such that at certain points in the spectrum their dispersion is a minimum. In optical glass this point lies in the far red. Borosilicate crown glass, of which the present prism is made, shows a difference in dispersion of nearly 40 times between 300 and 1000mμ. It was considered undesirable to have such a great variation in scale compression with wave-length.

The means selected to overcome the difficulty is essentially a sine-bar mechanism, as shown in Fig. 3. The wave-length scale is on the edge of a fan-shaped sector, turning on a pivot at its small end. Located on the fan is a drive-point connected by means of a spring-loaded link to a drive-point on the movable plate of the quadrilateral. The geometrical relationship of the drive-point on the fan, the drive-point on the quadrilateral, and the bearing of the fan is such that the three points move more and more nearly into a straight line as the red end of the spectrum and wave-length scale are approached. The result is a steadily decreasing rate of travel of the quadrilateral for a constant rate of rotation of the wave-length scale, giving a considerable expansion to the red end of the wave-length scale.

A wave-length range extending only to 320mμ in the violet permits the use of an incandescent light source throughout. In combination with a blue sensitive phototube, the energy available at 320mμ is sufficient to permit band widths of less than 5mμ. Between 320mμ and 500mμ there is a change of energy of 2500 to 1. To accommodate this there must either be a great range of gain control on the amplifier or the light intensity must be controllable by means of variable slits. Actually a combination of the two is ideal and is the mode of operation employed on the present instrument.

In order to set the reference or P_0 sensitivity[3] of an instrument to the arbitrary point required on the scale of a direct reading galvanometer, a continuously adjustable fine control, having a minimum of backlash and hysteresis, must be provided. On the other hand, a stepwise sensitivity control, giving accurate transmittance scale multiplying factors, has been found a great convenience, allowing all readings to be made on the low absorbance (high transmittance) end of the meter scale, thus permitting much increased analytical accuracy. The latter control cannot be optical because of considerations including spatial variation of absorption cell wall transmittance, photocathode sensitivity, and beam vignetting. To permit only two controls, then, the slit width must be continuously variable and must permit change of output energy in 0.1 percent or smaller increments. At widths of 0.05

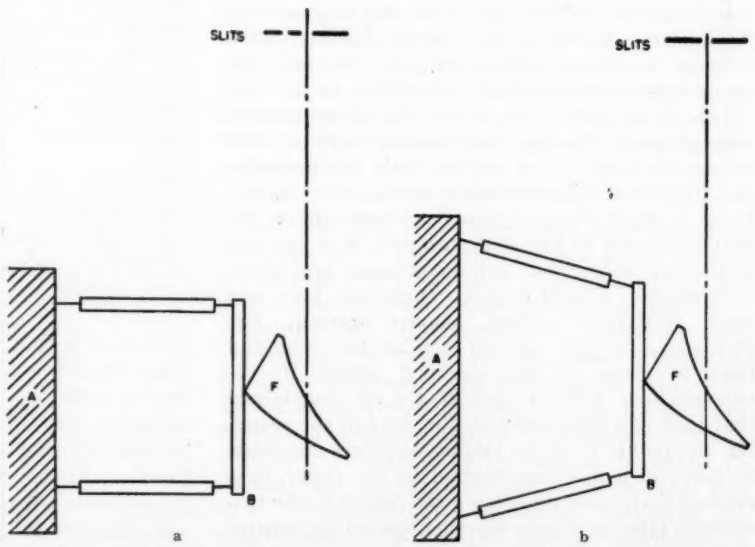

FIG. 2. (a) A parallelogram does not allow the prism to rotate, only to travel toward and away from the slits. (b) A quadrilateral causes the prism to rotate as it moves toward or away from the slit. When the relative separation of the hinges in the block *A* and the plate *B* are properly chosen, the rate of focal travel is coordinated with the rate of rotation of the prism.

[3] Committee on Colorimetry, Optical Society of America, "Physical concepts: Radiant energy and its Measurement," J. Opt. Soc. Am. **34**, 183–218 (1944); Nat. Bur. Stand., "Terminology and symbols for use in ultraviolet, visible, and infra-red absorptometry," letter circular LC857, May 19, 1947.

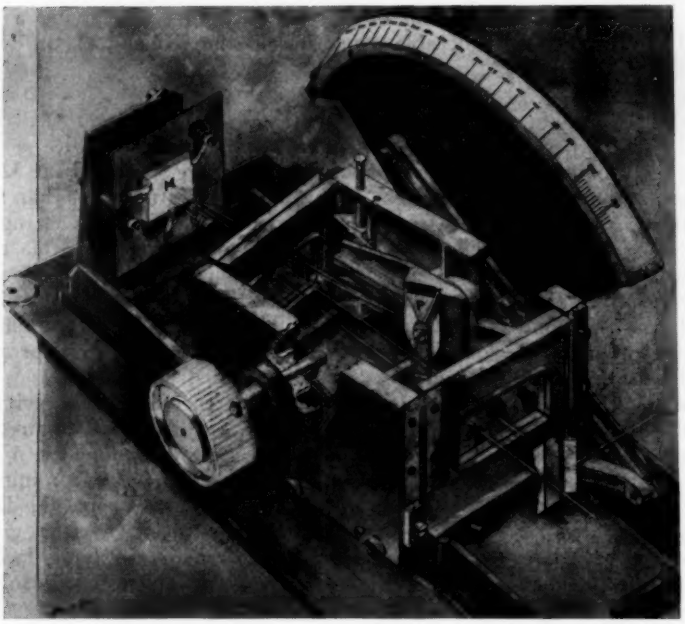

FIG. 3. A drawing showing the essential parts of the Fery prism monochromator. Light enters from the right, is deflected by a small 45° mirror through the curved entrance slit, proceeds to the optical flat *M*, thence to the Fery prism *P*, which disperses the light and returns it to the mirror *M*. It then passes through the straight exit slit to the sample compartment. Cut-away sections show the beryllium-copper reeds which act as hinges at the four corners of the quadrilateral, the gear reduction between the wave-length control knob and the wave-length scale, and the sine-bar linkage between the wave-length scale and the prism quadrilateral.

mm, this implies manual variation of geometric width in about 0.025 micron steps. By careful attention to design, this objective was achieved.

SLIT MECHANISM

Variable slits provide more than their fair share of problems. If bi-lateral slits are employed, it is necessary to move four slitjaws in perfect unison free from lost motion, backlash and any relative displacement between their centers. An alternative is to move only one jaw of each pair. Obviously the two selected must be those focused on each other.

Due to its greater simplicity, the latter method was adopted. Several mechanisms were studied before the problems of precise, wide range motion and freedom from hysteresis of motion were solved. From a single sheet of thin beryllium-copper, the pattern shown in Fig. 4 is stamped. *A* is one jaw of the exit slit, and *B* is the conjugate jaw of the entrance slit. Placed between these two jaws is a small stationary section (shown dotted). The two edges of this section act as the opposing jaws to those in the stamped plate. If the two working jaws *A* and *B* are so constrained that they can move only in the plane of the paper, but the points *C* of the joining members are made to move in a direction normal to the paper, it is evident that, upon depressing the points *C*, the two working jaws are drawn together toward the central stationary opposing jaws, closing them.

The members *D*, when bent up at right-angles to the plane of the paper, form four reed-type suspensions for the outer ends of the slit plate. Two more wide reeds, one fastened to each of the working jaws at the dotted lines *E* and *F*, provide suspensions for those members. Thus, the entire mechanism can be suspended on the ends of reeds free from any drag or friction with guides or ways.

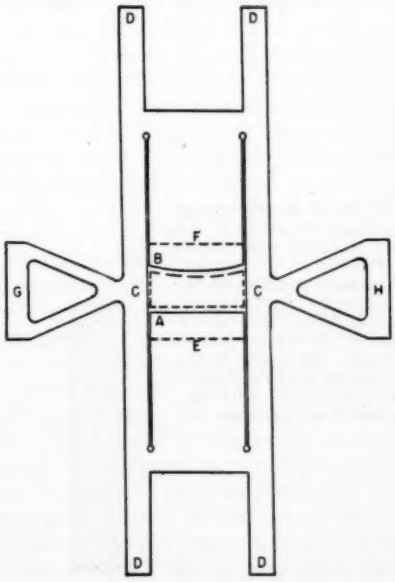

FIG. 4. Schematic drawing of the new slit plate described in the text, showing the curved entrance slit at *B* and the straight exit slit at *A*.

Suitable linkage between the points *C* and a micrometer provides a convenient means for their operation. Side members *G* and *H*, anchored at their ends, provide the centering restraint.

Figure 5 shows a perspective view of the complete slit mechanism and micrometer screw as employed in the spectrophotometer.

In addition to the advantage of avoiding the astigmatism of the prism, other considerations make a curved entrance and straight exit slit desirable. Narrower absorption cells can be used, and cylindrical cells give less trouble from exit beam deflections. Also, from considerations of chromaticity, a mirror condenser in the entrance optics is desirable. Such a mirror, inevitably working at high angles off-axis, produces a strongly curved image of the source. The source optics are designed to provide an image whose curvature agrees in direction with the entrance slit. Conditions of maximum optical efficiency demand that the tangential image of the source fall on the entrance slit and sagittal image fall on the entrance pupil of the monochromator (the prism). Fulfilling this condition exhausts the variables which could otherwise be employed to provide exact agreement between the curvature of image and the entrance slit. However, the image at the slit can be large enough to insure filling this aperture.

Variable slits, coupled with the dispersion characteristics of the glass prism, result in variable spectral band width of the light impinging on the sample. Considerable confusion exists in relation to accepted terminology for band width. Pending the general adoption of a standard system throughout the industry, this laboratory has accepted the following definitions, applicable to the optimum case where entrance and exit slits are of equal spectral width at all times, and where spectral variations of source energy and photoreceiver sensitivity are slight.

The *spectral slit width* is twice the mechanical slit width divided by the linear dispersion of the prism, at the wave-length of interest, in the plane of the exit slit. Neglecting scattering and aberrations, this includes all wave-lengths contributing any energy to the exit beam.

The *effective band width* (see Fig. 6) is the width of the band of wave-lengths in the exit beam, each of which contributes 50 percent or more of the energy of the nominal wave-length. This is equivalent to the half-energy band width for a line source. In practice, approximately 75 percent of the total energy in the exit beam lies within the wave-length range of the effective band width. Numerically, the effective band width is half the spectral slit width.

Incandescent light sources produce relatively little radiant energy in the violet. To obtain full scale deflections in the neighborhood of 320mμ, it is necessary to open the slits. As a result, the optical elements are flooded with light only a small fraction of which is energy of the desired wave-length. Since the exit slit opens along with the entrance slit, some stray light inevitably escapes along with energy of the nominal wave-length. To prevent this stray light from reaching the sample, the monochromator is provided with a Corning 9863 filter, for use from 320 to 400mμ. With this filter, stray light amounts to 1.3 percent at 320mμ when tested with a glass filter which cuts off all radiation of wave-length shorter than 340mμ. Tested with a filter which cuts off all radiation shorter than 360mμ, the stray light is only 0.2 percent. Stray light rapidly becomes negligible as wave-length is increased.

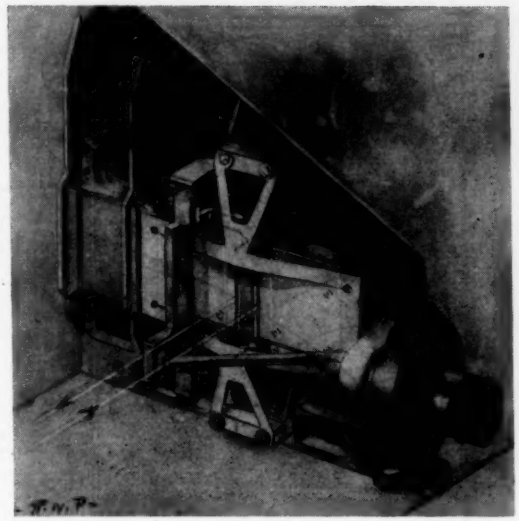

FIG. 5. A drawing showing the complete slit-plate assembly, micrometer screw, and supporting reeds.

AMPLIFIER DESIGN

The low light levels present in instruments of this type, plus the high order of accuracy and stability desired, pose special proglems in the design of the phototube and its associated amplifier.

To maintain high photoelectric sensitivity over the full spectral range of the instrument, two phototubes are employed. The red-sensitive tube, with an S-1 surface, may be used over the range from 400mμ to 1000mμ, and the blue-sensitive tube, with S-5 response, may be used from 320mμ to 700mμ without requiring excessive slit widths. The two phototubes have equivalent sensitivity at approximately 640mμ and may be interchanged conveniently in the instrument.

To achieve a high order of sensitivity, yet maintain excellent stability, a relatively low phototube

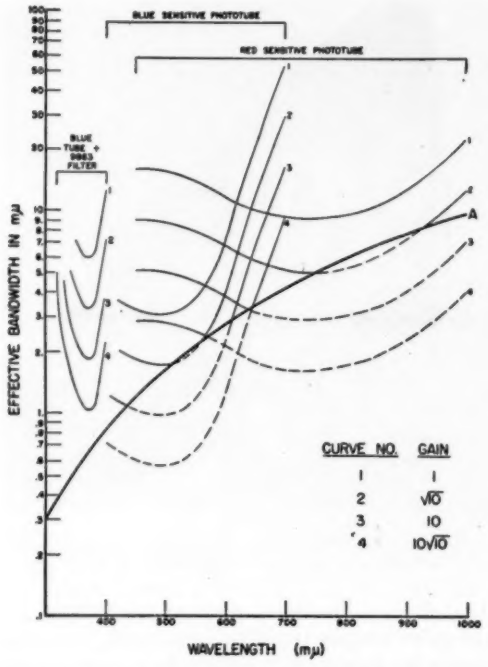

FIG. 6. Typical effective band widths obtained with the Fery prism spectrophotometer are shown for the entire wavelength range of the instrument, at all of the four gain-selector switch positions, for the two phototubes (red- and blue-sensitive) used with the instrument. The heavy curve *A* gives typical effective band widths for the instrument when a constant slit width of 0.1 mm is maintained.

load resistor is employed (500 megohms) in conjunction with an exceptionally stable high gain d.c. amplifier utilizing the latest standard types of miniature and sub-miniature tubes. This combination makes possible unusual stability, freedom from drift, rapid response, and phototube current sensitivity exceeding 5×10^{-13} ampere.

Accuracy, stability, and rapid response of the amplifier are further assured by the use of negative feedback, an electronically regulated voltage supply, and a high torque, highly linear meter with mirror scale to avoid parallax.

The electronic voltage regulator secures its reference voltage from a divider network which takes advantage of the non-linear current-voltage relationship of Thyrite.[4] The regulator provides supply voltages which are stable to one part in ten-thousand for a ten percent line voltage change, which makes it possible to operate the amplifier directly on the 115-volt line without providing additional line voltage stabilization. With this degree of regulation, the meter reading changes less than 0.1 percent for a 10 percent line voltage change.

Other features that contribute to the unusual stability of the instrument include shock-mounting of all critical elements, thermal isolation of the Thyrite and phototubes, light shielding of the input amplifier to eliminate photoelectric effects, parallel switch contacts in low impedance circuits for increased reliability, and specially selected resistors in the input circuit.

Convenient adjustment has been assured by incorporating a 10-turn Helipot as the dark-current control, which accommodates a wide range of individual differences in phototubes and amplifier input tube bias, yet provides a sufficiently fine adjustment when operating on the most sensitive range.

The amplifier sensitivity is controlled by a 4-position switch, selecting precision resistors in the feedback network, each step representing a factor of $(10)^{\frac{1}{2}}$. Transmission measurements are multiplied directly by these factors while absorbance readings have 0.5, 1.0, and 1.5 added, respectively, for the three steps. This feature makes precise readings possible on samples with high optical absorbance and permits routine analysis to be made more accurately using lower sensitivity and increased band width when this is desirable. In addition, by suitable selection of working concentrations, all transmittancy readings can be made at nearly the

TABLE I. Beer's law tests on potassium chromate in 0.05 *N* KOH at 372mμ.

K_2CrO_4	100 ppm	50 ppm	25 ppm	12.5 ppm	6.25 ppm	3.125 ppm
Transmittancy (I)	0.8380	0.7016	0.4927	0.2417	0.0600	0.00394
Absorbancy ($\log I_0/I$)	0.0767	0.1539	0.3074	0.6176	1.2217	2.4048
Mole. conc.	1.61×10^{-5}	3.22×10^{-5}	6.45×10^{-5}	12.9×10^{-5}	25.8×10^{-5}	51.5×10^{-5}
E	4766.83	4780.00	4765.75	4780.75	4735.24	4669.58
Wt'd. mean E	4753.52					
Computed transmittancy (I')	0.83843	0.70297	0.49363	0.24367	0.05937	0.003564
$I'-I$	+0.00043	+0.00138	+0.0009	+0.00196	−0.00063	−0.00038

[4] General Electric Descriptive Bulletin GEA-4138A.

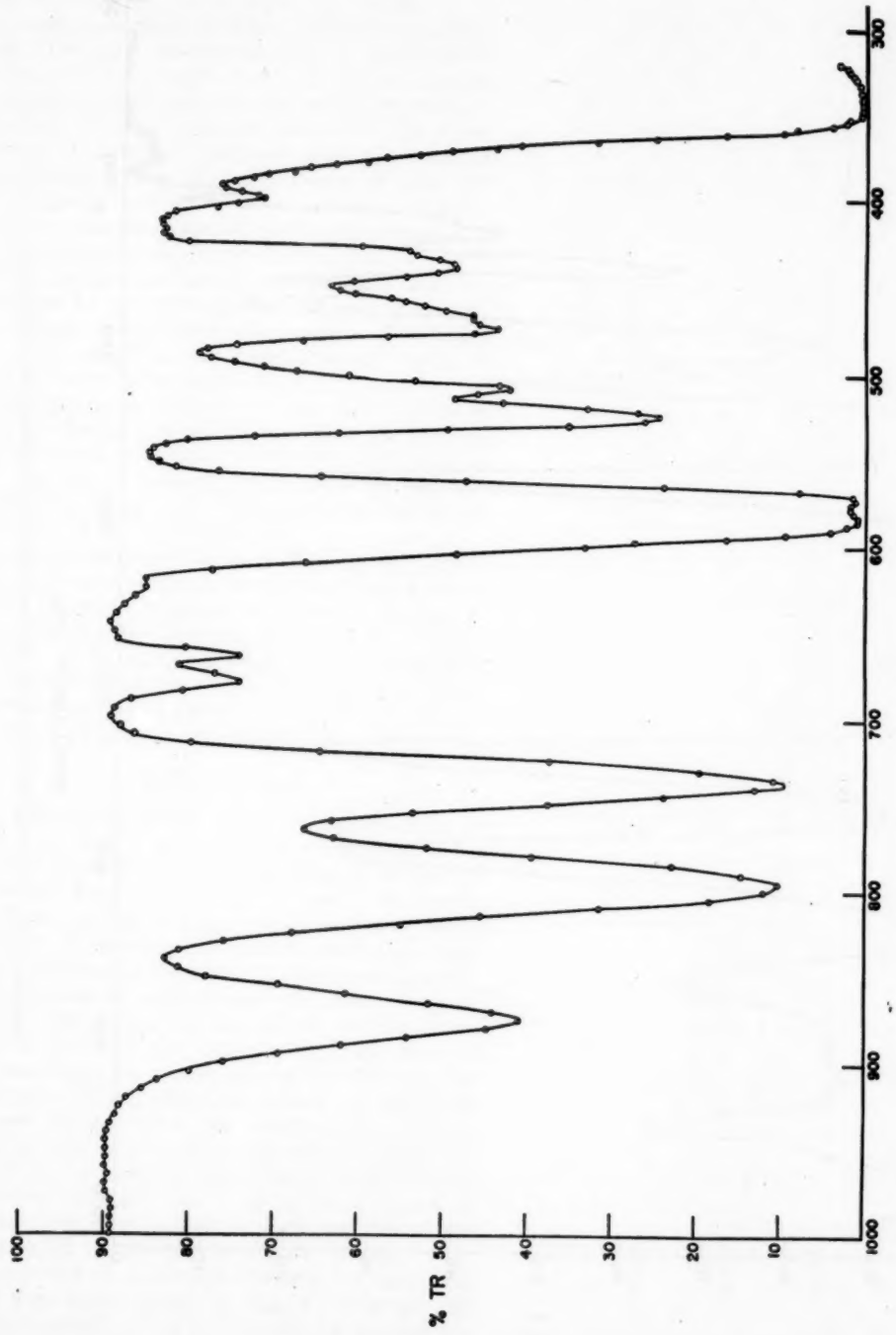

FIG. 7. Transmission curve of a didymium glass filter determined in the manner described in the text to show up any lack of reproducibility of the wave-length mechanism.

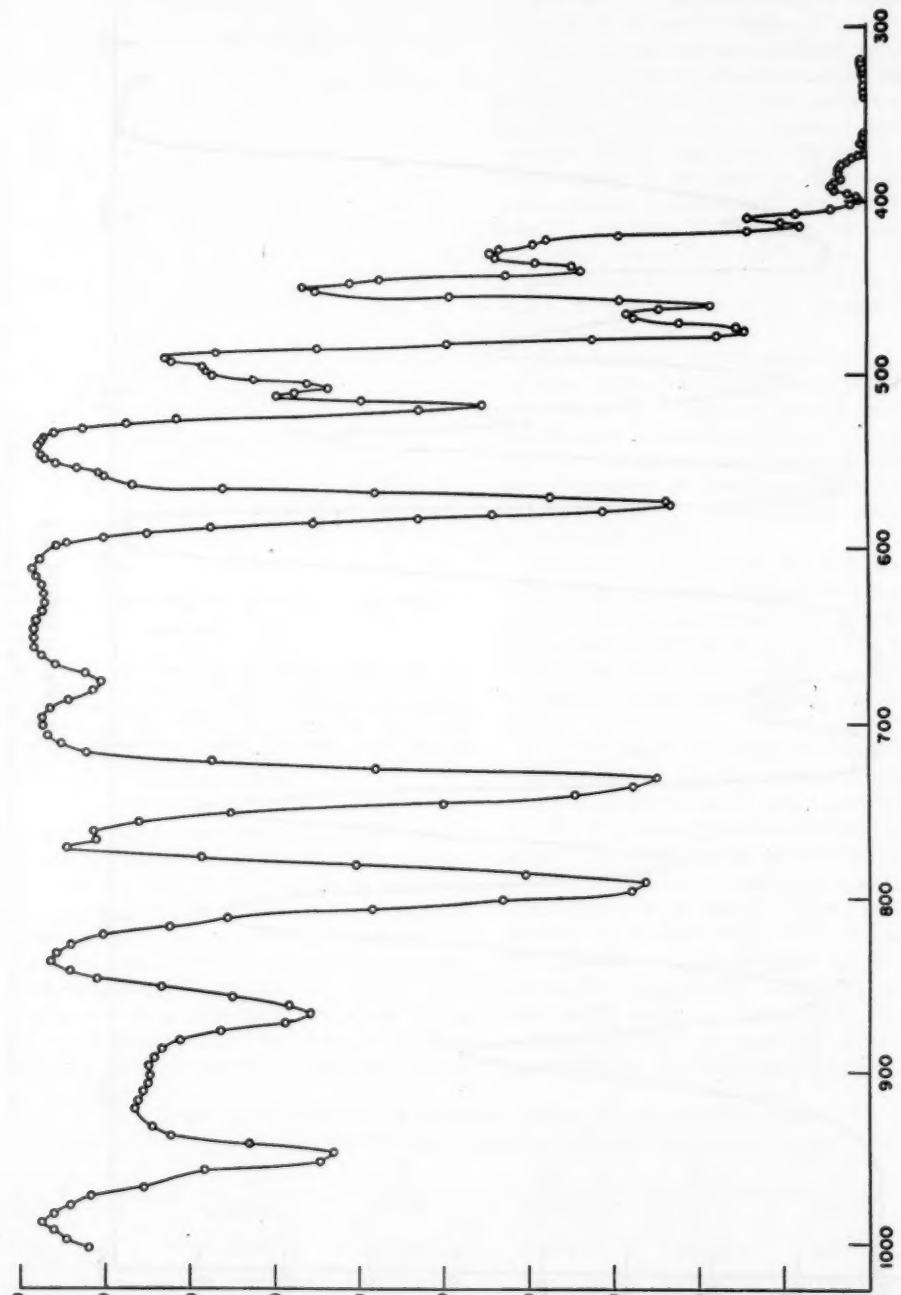

FIG. 8. Transmission of a samarium chloride solution measured with the Fery prism spectrophotometer.

same meter deflection, effectively compensating the small non-linearity errors of the meter scale.

Operating controls have been reduced to a minimum. The gain selector switch, a dark-current control, the wave-length knob, slit control, sample placement control, and the shutter and filter knob are all conveniently available along the front of the instrument. Normal operation sequence requires wave-length setting, slit adjustment to give 100 percent scale reading on the standard, and insertion of the sample to permit reading the absorbancy or transmittancy directly from the meter.

The carriage bearing the photocell and load resistor can be quickly removed, not only to permit the change to the red-sensitive photocell, but also to permit access to the monochromatic beam of light for use in other applications.

The lamphouse with its 6-volts, 32-cp prefocus bulb can be opened giving access to the entrant light beam in those cases where analysis of special light sources is desired. An example of such an application is the now widely accepted analytical technique of flame photometry.

Installations providing stable line voltage permit the incandescent lamp to be operated from a transformer located in the amplifier section of the instrument. Suitable terminals are provided. For utmost stability, or in locations where the line voltage is not sufficiently stable, a storage battery or other controlled source of power must be employed. An a.c.-operated lamp power supply is being developed.

PERFORMANCE

A spectrophotometer must be reproducible in both wave-length and photometric accuracy. To test the wave-length reproducibility, a sample of didymium filter glass was measured. The results are shown in Fig. 7. The method of measurement consisted of determining the transmittance every 100mμ while going from short to long wave-lengths. The direction of measurement was then reversed, and the transmittance determined for the intermediate 50-mμ points. Again going from short to long wave-lengths, measurements were made at the intervening 25-mμ points. In a final run from long to short wave-lengths, the remaining points were filled in. Any lost motion, hysteresis or other defect likely to result in lack of wave-length reproducibility, will be shown up in this manner of measurement, particularly at such portions of the curve as the steep slopes at 375 and 600mμ.

A solution of samarium chloride was measured with the results shown in Fig. 8. Although the structure is not as fine as it should be to show the resolution of which the instrument is capable, no other materials with finer structure in the visible spectrum are known to us.

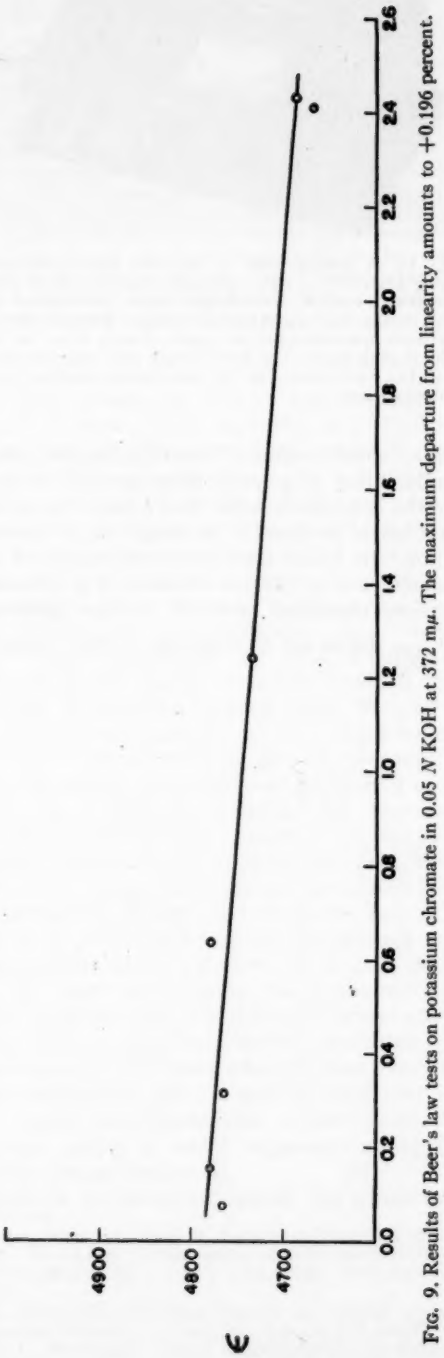

FIG. 9. Results of Beer's law tests on potassium chromate in 0.05 *N* KOH at 372 mμ. The maximum departure from linearity amounts to +0.196 percent.

FIG. 10. A general view of the new spectrophotometer, showing from left to right, the gain control selector switch, dark-current control, wave-length knob, slit control knob, sample compartment, and shutter control. Transmittance and absorbance measurements are made directly from the meter on the sloping panel. The wave-length scale appears directly above. The switches at the left end control amplifier current and light source.

An accepted method of checking the photometric reproducibility of a spectrophotometer is to determine the compliance with Beer's law of a series of transmission or density readings on a chromate solution.[5] In Table I are given the results of such measurements on various dilutions of a solution of potassium chromate in 0.05*N* sodium hydroxide.

[5] H. von Halban and K. Siebentopf, J. Phys. Chem. **100**, 208 (1922).

The values labeled $I'-I$ are the differences between the transmittancy computed from the mean value of E and the measured transmittancy of each sample. It should be noted that the maximum errro amounts to less than 0.20 percent, fully justifying the claims of 0.50 percent accuracy made for this instrument. The results are also shown in the graph in Fig. 9.

Figure 10 shows a view of the instrument as a whole. The inclined panel permits operation from either a standing or a seated position. The wavelength and absorbance dials are placed adjacent to each other to permit rapid measurements. The instrument is portable, weighing 47 lbs. The sample compartment accommodates several types of accessories such as cell carriages for 1-cm square cells, tubular cells up to 50 mm in length, standard test tubes, an integrating-sphere reflectance attachment, or other devices required by special problems.

ACKNOWLEDGMENT

Many individuals assisted in the design of this instrument. Special credit is due E. R. Schweitzer, R. R. Davis, and J. E. Jacobs for valuable aid in design details and testing. The work was carried out under the supervision of R. C. Hawes, research director of these laboratories.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 5 MAY, 1949

Comparison of Color Systems with Respect to Uniform Visual Spacing

ROBERT W. BURNHAM
Color Control Department, Eastman Kodak Company, Rochester, New York
(Received February 2, 1949)

The present study is a comparison of ten systems of color specification, for a restricted chromatic area around the neutral at a single luminance level, with respect to their relative adherence to uniform visual spacing. A sample of 20 visually equi-spaced (Munsell renotation) chromaticities at the same luminance were used. Chromatic coordinates for these colors were expressed in each of the ten systems and plotted in their respective chromaticity diagrams. Then all data were reduced to the same scale, after which radial saturation distances and hue angles were compared to those in the Munsell renotation criterion. Rank orders of departure from the criterion were determined for hue and saturation separately. Results show that all the 10 systems include visually significant departures from a uniform color space. These results may be generalized beyond the present sample because, if other more saturated colors were added to the present sample, there would be even greater departures from uniform spacing. The use of any particular system in any particular instance depends on circumstances, the time and personnel available, the number of points to be transformed, and the accuracy with which it is desired to illustrate visual relationships. Where the accurate expression of visual relationships is of prime importance, Munsell renotation is recommended because it was designed specifically to show equal visual intervals.

I. INTRODUCTION

IT is well known that a given distance between two colors in one part of the I.C.I. chromaticity diagram is not equal visually to the same distance in other parts of the diagram. Because the I.C.I. system was designed to express specific psychophysical relationships, it is not unreasonable to expect that psychodynamic[1] visual relationships may be only incidentally and imperfectly expressed. That this is the situation can be shown when an attempt is made to use the I.C.I. system for representing visual relationships among colors, i.e., to use it as a direct measure of color differences in terms of the observer's awareness of them.

If one's interest lies principally in a unique specification for colors, such an inhomogeneity is not important; the affine, non-metric characteristics of the I.C.I. system are adequate without alteration or transformation. When, however, there is special interest in the magnitude and uniform expression of color differences, it becomes important either (1) to modify the I.C.I. system so that an homogeneous, isotropic color space results in which equal angles and distances have equal (or approximately equal) visual meaning, or (2) to turn to some other notation explicitly designed to express these psychodynamic relationships.

In line with the first of these two possibilities, *viz.*, modification of the I.C.I. system, attempts have been made by Judd; Breckenridge and Schaub; Scofield, Judd, and Hunter; MacAdam; Adams; Moon and Spencer; Farnsworth; Saunderson and Milner;[2] and others, to transform the I.C.I. system into one which includes uniform color spacing. The emphasis in these studies has been to give psychodynamic meaning to an existing and widely used psychophysical system. Silberstein[3] has, incidentally, given rigorous mathematical proof, using MacAdam's small-difference data,[4] that no linear transformation of the I.C.I. diagram can represent exact equality of small (liminal) visual intervals. In the face of such proof, Wright[5] has pointed out that, considering the limited accuracy with which visual intervals can be measured, the variation among observers, and the varied uses to which a uniform scale would be put, there is no need to demand exact equality over the whole diagram. For his own small-difference data, Wright showed, by using the Breckenridge-Schaub RUCS transformation of the I.C.I. diagram, how it was possible to reduce the ratio of the shortest to the longest of his I.C.I. difference steps from 20:1 to 4:1, and how—by ignoring the purple-red corner of the diagram—the ratio was still further reduced to 2:1. Wright stated, in addition, that other transformations may be found which give an even closer approximation to uniform spacing, and that there is no reason to suppose that a new chromaticity diagram giving a useful improvement in color spacing cannot be derived.

There is, as mentioned above, the second possi-

[1] The term, psychodynamic, has been defined by Guilford to contrast with the term, psychophysical, as follows: "in a psychophysical problem we relate measurements made on some psychological continuum to measurements made on some corresponding physical continuum. In a problem in psychodynamics we relate two sets of measurements both made on certain psychological continua · · · our interest is confined to responses and the relations between them." J. P. Guilford, "A study in psychodynamics," Psychometrika 4, 1–2 (1939). The psychodynamic relationship referred to in the present study is that between pairs of colors which are visually equidistant. Systematic expression of equal visual distances can certainly be found in a color space which is designed to express such equality.

[2] Basic references to all of these transformations are given in: D. Nickerson, "Interrelation of color specifications" in "A Symposium on Color." Tech. Assn. Papers, 472–490 (1947).

[3] L. Silberstein, "Investigations of the intrinsic properties of the color domain," J. Opt. Soc. Am. 33, 1–9 (1943).

[4] D. L. MacAdam, "Visual sensitivities to color differences in daylight," J. Opt. Soc. Am. 32, 247–274 (1942).

[5] W. D. Wright, *The Measurement of Colour* (Adam Hilger, Ltd., London, 1944), pp. 182–188.

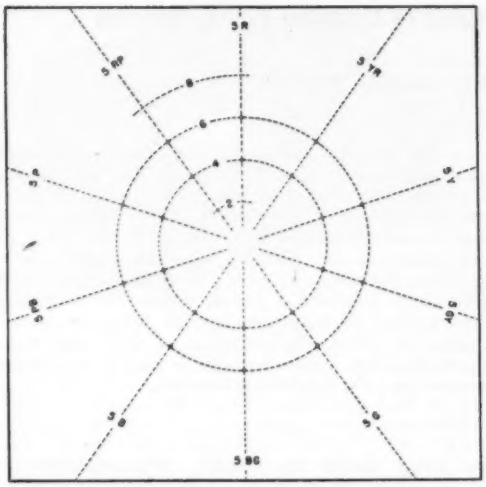

FIG. 1. Munsell plot of test color samples.

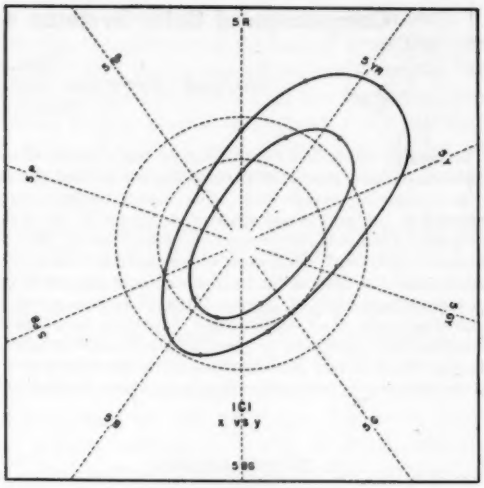

FIG. 3.

bility of designing a notation for the systematic expression of psychodynamic visual relationships with only secondary concern for the psychophysical relationships inherent in the I.C.I. system. This approach, too, has been taken. The O.S.A. Subcommittee report on renotation of the Munsell colors[6] embodies an elaborate empirical attempt to specify colors by means of scales that have visual meaning.

In an experiment by the writer pertaining to tolerances for photographic color prints it became important to express I.C.I. color difference data in visually meaningful terms. Of particular interest were allowable chromatic departures from neutral. The same need for a uniform color space has also arisen in other industrial situations where the color of a product is a sizeable factor in its appeal to the consumer. It is, in fact, almost impossible to work with color specifications as a set of numbers and symbols without running into the need, at least occasionally, for knowing what the specifications represent visually. Then, too, inaccurate data and publication of misleading results have, as Nickerson[7] states, been avoided more than once by knowing the visual meaning of color specifications.

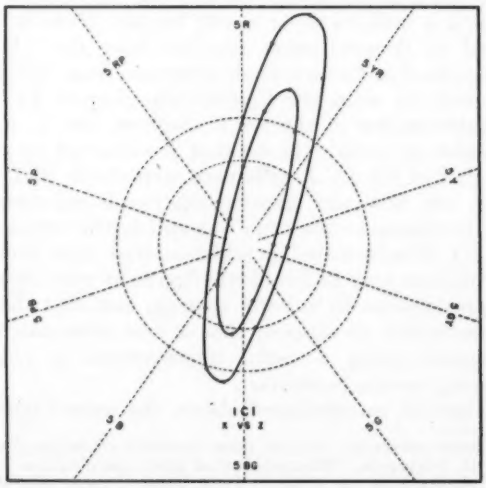

FIG. 2. Comparative plots of 10 systems of color specification with respect to the Munsell renotation uniform color space are shown in Figs. 2–11.

Munsell renotation scale distances were designed to correspond to observed color differences; equal distances on each of the hue, chroma, and value scales were made to represent equal observed visual intervals of that attribute. Since no other known color system is based on such a correlation, the probability is great that Munsell scales provide the best available approximation to uniform visual spacing. Variability among and within the observers and the smoothing techniques used in the renotation procedure prevented the achievement of perfect uniform spacing, but the use of voluminous observational data has resulted in the best approximation yet produced.

Munsell renotation appeared to the writer to be the system in which the experimental data, referred to above, should be expressed in order to show visual relationships. Transformation of I.C.I. data to Munsell renotation involves, however, inconvenient interpolations.[6] For this reason, and

[6] Newhall, Nickerson, and Judd, "Final report of the O.S.A. subcommittee on the spacing of the Munsell colors," J. Opt. Soc. Am. **33**, 385–418 (1943).

[7] D. Nickerson, "The Munsell color system," Illum. Eng. **41**, 558 (1946).

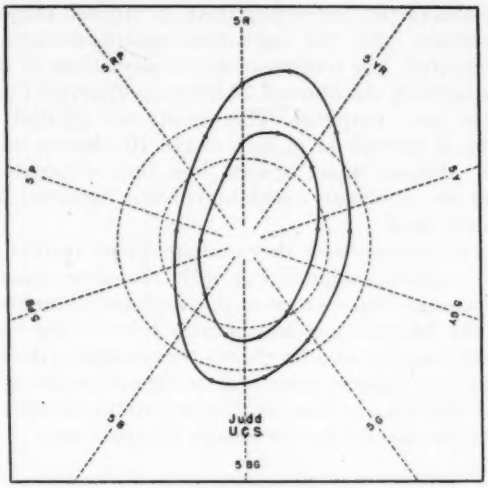

FIG. 4.

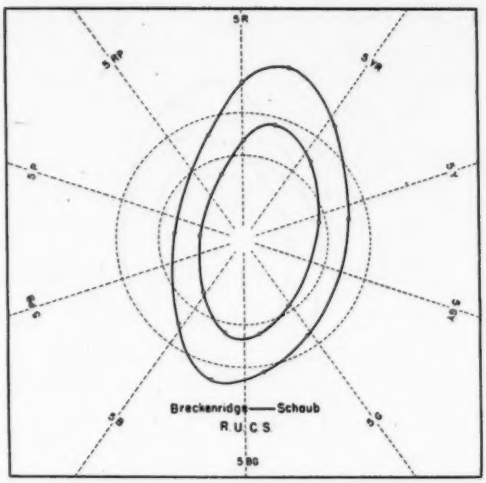

FIG. 5.

because of the utility and broad usefulness of the I.C.I. system in other connections, the question arose as to how near to this "empirical ideal" the various color systems come which are mathematically direct transformations of the I.C.I. system. If any particular system departs from the empirical ideal by less than visually significant differences, then the mathematical economies involved (adaptability to machine computation), and the more direct transformational relationship to fundamental measurements inherent in I.C.I. specifications, should render such a system frequently more practical than the Munsell renotation for the expression of uniform visual intervals.

The problem of the present study has, therefore, evolved into a determination of the extent to which ten of the better-known color systems depart from the Munsell renotation criterion in uniformity of color spacing.[8] These include the I.C.I. system and various other systems which are transformations thereof.

II. PROCEDURE

The ideal uniformity found in graphic Munsell plots of 20 visually equi-spaced colors was compared with variations therefrom in the ten selected systems of color specification, in order to determine which, if any, of the systems approximate the Munsell criterion by less than visually significant variations, and to find out how the various systems compare with one another with respect to the sample of colors used in this investigation.

[8] The study was undertaken in full recognition of the fact that different systems have been developed to fill specific requirements and to be useful under given conditions. Each author has, however, stated one of his aims to be the achievement of uniform scaling of color space.

Ten hues were selected at equal intervals around the renotated Munsell hue circle, as shown in Fig. 1, beginning with 5*R*, at a renotation 6 value and at each of the renotation 4 and 6 chroma distances from neutral, making 20 colors in the test sample (10 hues at 4 chroma and 10 hues at 6 chroma). On a Munsell color plot, the angles between adjacent hues were all equal (36°) and each represented 10 Munsell hue steps, and the radial chroma distances from neutral were equal for all hues at each of the 4 and 6 chroma levels.

The equal 36° angles between adjacent hues, and the equal chroma radii from neutral to the samples at each chroma locus, served as the two criteria against which the various systems were compared for uniformity of spacing.

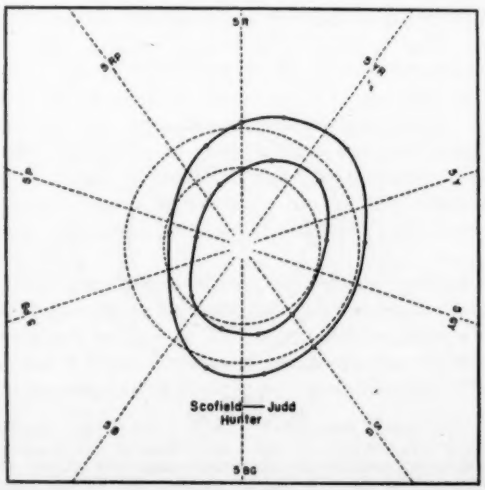

FIG. 6.

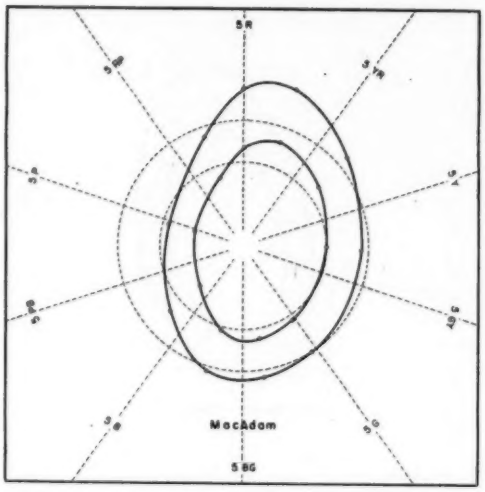

FIG. 7.

In order to compare the various systems, the renotation specifications for each of the 20 sample colors were transformed to I.C.I. coördinates (based on Illuminant *C*) and were then transformed to the coördinates of each of the other systems. Then, taking the Illuminant *C* point as the origin, the rectangular coördinates of each sample point were transformed to polar coördinates. The mean radial distance for the 20 sample colors was determined for each system and the factor necessary to convert each mean radius to Munsell chroma 5 found. (Munsell chroma 5 was the mean chroma of the 4 and 6 chroma sample colors, and it was desired to convert all of the 10 systems to the same scale as the Munsell criterion.) The radii were multiplied by this factor, converting them to Munsell chroma numbers, and the root-mean-square deviations computed. The root-mean-square deviations of the angles from the Munsell 10 hue-step intervals (36°) were also computed. The results are plotted in Figs. 2 through 11 in each of the 10 systems with the Munsell equal spaced hue lines superposed and the circles of constant chroma indicated by dotted lines.

The systems were then compared with respect to the relative magnitude of their root-mean-square deviations (square root of the variance) from each of the Munsell hue and chroma criteria, the idea being that the smaller the root-mean-square deviation, the closer a particular system approximated the criterion—at least with respect to the chromatic area represented by the sample of colors used.

III. RESULTS

The general result was that none of the systems approximated a uniform color space by less than visually significant root-mean-square deviations.

Root-mean-square deviations for hue correlated with those for saturation by a Pearson product moment correlation coefficient of 0.86, showing a near 1 to 1 correspondence between the relative values of the hue deviations as compared to the chroma deviations for each system.

Figure 3 shows the basic I.C.I. x, y plot. It will be apparent that the departure of the I.C.I. points from ideal circularity is great; angular differences and differences in radial length are obvious. It is no wonder that for visual color work it was decided to look for some transformation of the I.C.I. system embodying more uniform spacing.

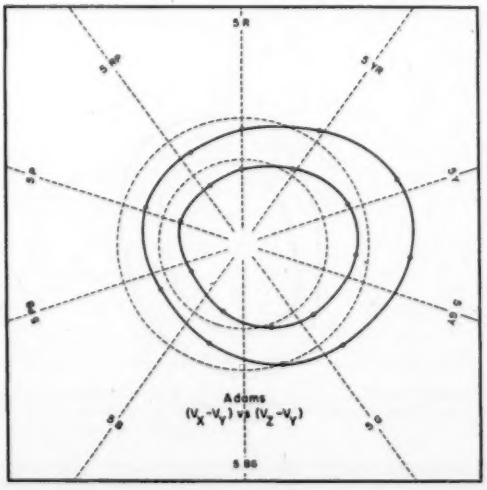

FIG. 8.

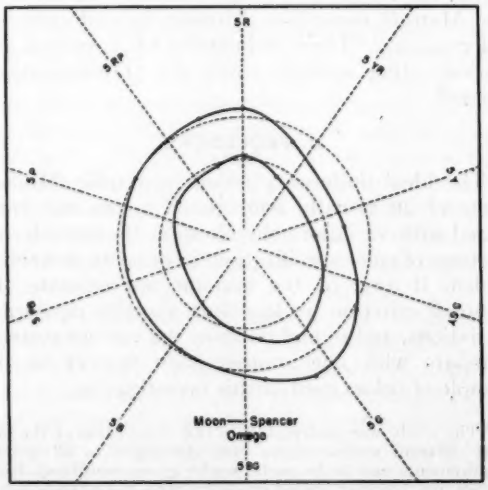

FIG. 9.

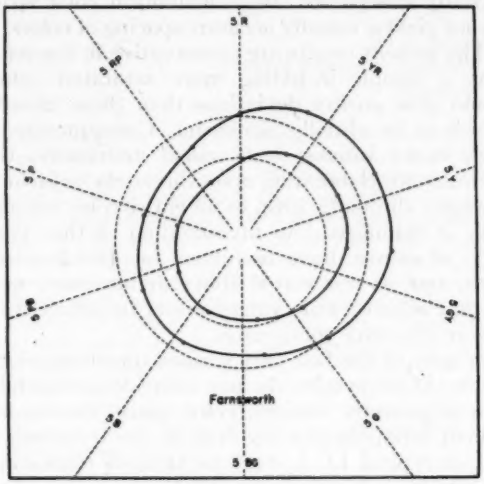

FIG. 10.

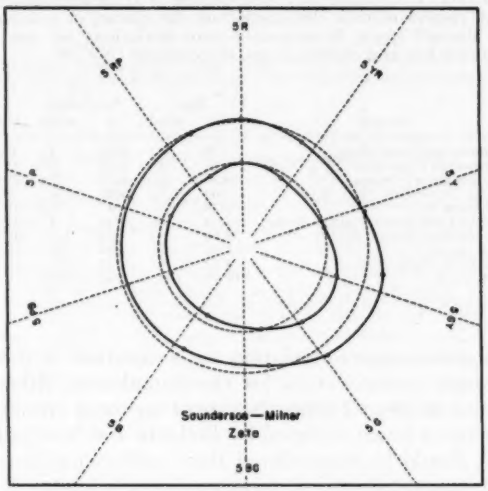

FIG. 11.

Figure 2 shows a plot of the I.C.I. x, z coordinates of the 20 colors. The x, z plot is one recommended by Adams because, as he states,[9] "Y is the unique coördinate to which all luminosity of the I.C.I. system is attached and an x, z plot would, therefore, seem more logical than the x, y plot." In spite of the logic of such a plot, it so happens that there is a greater departure from uniform chromaticity spacing in the x, z plot than in the x, y plot.

It is obvious from the comparison plots that some are more nearly circular and uniform than others. Table I summarizes the computations on which the color plots were based. Ranks of the systems are shown in parentheses in Table I for hue and chroma separately. The order of the systems in Table I is based on the mean of hue and saturation ranks. Hue and saturation root-mean-square deviations are expressed in Munsell hue and chroma steps, respectively, and represent relative departures of the actual angles and radial distances from the Munsell criterion.

Chronology of the various systems is shown in the last column of Table I since there is an encouraging indication of progressively closer approximation to uniform spacing.

Because of the limited sample, no attempt was made to combine scale values or ranks to provide a single scale value for each system. It may be of some interest to state that hue and chroma steps may be converted to the same unit, the difference limen (DL). One hue step equals 1.8 DL and 1 chroma step equals 5.1 DL.[10]

IV. DISCUSSION AND CONCLUSION

The present results are based on a sample of colors in a limited chromatic area at a single luminance level (Munsell 6 value). To extend the implications of the relative scale values beyond the sample would be dangerous and unjustified. Any other comparable sample of colors, or a larger sample would very likely produce different relationships among the systems. The present relative scale values apply only to the sample involved. If certain available facts are considered, however, the results do provide a basis for generalization in at least one respect. For any of the transformations of the I.C.I. system, minimal departures from uniform spacing are to be found in the area immediately surrounding the neutral point at any given luminosity level. Greater than minimal departures are obvious when an inspection is made of color plots in more saturated areas of the various transformations. If departures from uniform spacing in the near-neutral areas are of visual significance—i.e., if the departure is greater than just-noticeable—then it can be expected that departures of even greater visual significance would be found in more saturated areas of the several chromaticity diagrams.

The Saunderson-Milner zeta-space, one of the two top-ranking systems, shows a root-mean-square deviation for hue of 2.19 Munsell hue steps which is over 4 times the size of the mean hue difference limen reported by Bellamy and Newhall;[10] the

[9] E. Q. Adams, "X–Z planes in the 1931 I.C.I. system of colorimetry," J. Opt. Soc. Am. **32**, 168–173 (1942).

[10] The values for DL's given here are based on the mean of the colors, nearest to the chromatic area represented by the color samples used in the present study, as reported by B. Bellamy, and S. M. Newhall, "Attributive limens in selected regions of the Munsell color solid," J. Opt. Soc. Am. **32**, 465–473 (1942). The mean limens were determined from reported limens for 5*R* 5/6, 5*R* 5/2, 5*Y* 5/6, 5*Y* 5/2, 5*G* 5/6, 5*G* 5/2, 5*B* 5/6, 5*B* 5/2, 5*P* 5/6, 5*P* 5/2.

TABLE I. Comparison of 10 systems of color specification with respect to their deviation from the spacing of a sample of Munsell space. Root-mean-square deviations (σ) are in Munsell hue and chroma steps, respectively ($N=20$).

System	Hue		Saturation		Date
	σ	Rank	σ	Rank	
Saunderson-Milner (Zeta)	2.19	(3)	0.36	(1)	1946
Farnsworth Projection (I.C.I. x, y)	2.04	(1)	0.79	(3)	1944
Moon-Spencer (Omega)	3.37	(4)	0.59	(2)	1943
Adams (V_X-V_Y), (V_Z-V_Y)	2.04	(1)	1.08	(6)	1942
MacAdam (u, v)	4.28	(5)	1.02	(5)	1937
Scofield-Judd-Hunter (alpha, beta)	4.33	(6)	0.93	(4)	1941
Breckenridge-Schaub RUCS (x'', y'')	5.68	(8)	1.23	(7.5)	1939
I.C.I. (x, y)	4.93	(7)	1.93	(9)	1931
Judd UCS (R, G, B)	5.72	(9)	1.23	(7.5)	1934
I.C.I. (x, z)	10.99	(10)	2.64	(10)	1931

root-mean-square deviation for saturation of 0.36 Munsell chroma steps for the Saunderson-Milner system is about 2 times the size of the mean chroma difference limen reported by Bellamy and Newhall.

It should be remembered that a difference limen is a statistical entity, representing a stimulus interval that is perceived as different 50 percent of the time, and is, therefore, only about $\frac{1}{3}$ the size of a visually just-noticeable-difference (JND) which can be perceived 100 percent of the time. Even at that, it may still be said by way of interpretation that variation in one of the best known approximations to Munsell uniform color space (Saunderson-Milner zeta-space) may be from 1 to 6 times the size of a JND. This statement is based on the fact that virtually 100 percent of the total variation is included in 6 times the root-mean-square deviation, and since 1 root-mean-square deviation includes $\frac{2}{3}$ to 1 JND, then the greatest variation from uniform spacing may be as large as 6 JND's for a plot of the colors in the present sample (and even larger for samples including more saturated colors).

This reasoning is intended to substantiate the adequacy of the sample used in the present study and to provide the necessary basis for a conclusion that the best known approximations to the empirically determined Munsell uniform color space do not yield a visually accurate spacing of colors.

The present results are conservative in the sense that a sample including more saturated colors would give greater deviations than those already shown to be visually significant. Consequently, if one's major interest is in visual uniformity, the necessity for elaborating upon this study to include a greater chromatic area, or an entire color population, is minimized. An investigation of that type may, of course, have usefulness in other connections, and, as was stated above, it may very well be that relative scale values would be different in a more extensive comparison.

In spite of the fact that various transformations of the I.C.I. system do not coincide completely with a perfectly uniform color space, the inconvenient interpolations involved in the transformation of typical I.C.I. data to Munsell renotation may make it highly impractical to use the more uniform system. Most of the transformations are improvements on the basic I.C.I. system so far as uniform spacing is concerned, and the use of any one system might be recommended depending on particular circumstances, the time and personnel available, the number of points to be transformed, and the relative accuracy with which it is desired to illustrate visual relationships. On the other hand, where visual relationships are of prime importance, it is well to understand that—inconvenient or not—there is no comparable substitute at present for the Munsell space. This fact should not be surprising if it is remembered that the Munsell diagram was designed to fit observational data. Even the Munsell representation is limited, however, by the approximate nature of the observational data and the smoothing procedures, but since no other systematic set of observations has yet approached the experimental breadth of the Munsell renotation there is at present no other available criterion.

Transmission of Light from Diffuse Sources by Periscope Tubes

F. P. BUNDY AND H. M. STRONG
Research Laboratory, General Electric Company, Schenectady, New York

(Received February 3, 1949)

This paper treats in detail the problem of transmitting light from a diffuse source, such as a flame, to a place some distance away for the purpose of making spectroscopic measurements. The particular cases which have been analyzed are: the simple uniform tube, the simple tapered tube, the uniform tube containing one lens placed at its focal distance from the light source, the uniform tube with one lens of focal length equal to one-fourth the length of the tube placed at the midpoint of the tube, the uniform tube with a series of lenses spaced at intervals of twice their focal length, and uniform simple tubes with reflecting walls. In each case expressions for the total amount of light transmitted have been derived and the relative strength of the contribution from different zones of the source have been investigated. The greater part of the work is theoretical but most of the conclusions have been tested experimentally.

It is found that a tube with reflecting walls and no lenses gives the greatest amount of transmitted light. This light emerges from the output end of the tube nearly parallel to the axis but is disordered enough that the formation of sharp images is impossible. However, this is usually satisfactory in transmitting light to instruments such as interferometers and spectrographs when sharp images are not necessary.

A tube fitted with lenses spaced two focal lengths apart gives efficient transmission of light as well as formation of sharp images of the source at intervals of four focal lengths along the tube. Thus if it is desired to use light from specific parts of the diffuse source the "2F" lens system should be used even though it transmits less light than the open tubes with reflecting walls.

INTRODUCTION

IN making spectroscopic studies of the light emitted from rough and dangerous flames, such as those from rocket and ramjet motors, it is often necessary to transmit the light a considerable distance from the flame to a place where the optical equipment and the observer can operate satisfactorily and safely. Due to the fact that spectrographs and interferometers waste a large fraction of the light incident upon them it is often difficult to get satisfactory photographic exposures in the duration of a flame test run. Consequently, it is necessary to transmit the greatest possible amount of light from the flame to the optical instrument in such a manner that the light is used most efficiently by the instrument. In the present paper the emphasis is placed on the problem of transmitting the light from the flame to the output end of the periscope tube. It is also important that the optical instrument be arranged to make efficient use of the light transmitted by the periscope tube but this will not be given much consideration here.

In the theoretical analyses which follow it will be assumed that the flame is replaced by a plane diffuse source of uniform brightness of B units of power per unit of area and solid angle, that the periscope is of radius "a" and extends the full distance L from the light source to the output end, that the tube is perfectly straight and that its inside walls do not reflect light (unless otherwise stated), that any lenses placed in the tube have the full diameter of the periscope tube, and that the ratio of the length to the radius of the tube is large.

UNIFORM TUBE WITHOUT LENSES

The solid angle subtended by the output end of the tube as seen from the source plane at the input end is $\pi a^2/L^2$. Thus the contribution of each element of area dA on the source plane is

$$dQ = B(\pi a^2/L^2)dA \tag{1}$$

and hence the total light flux is simply,

$$Q_0 = (\pi a^2 B/L^2)\int_{\text{source}} dA = (\pi^2 a^4/L^2)B. \tag{2}$$

Thus the light flux increases as the square of the ratio of the cross-sectional area to the length. This light flux, Q_0, given by a plain tube with non-reflecting walls will be used in the remainder of this paper as the unit of light flux to which that of other systems will be compared.

TAPERED TUBE WITHOUT LENSES

In this case let A_1 denote the area of the input end of the tube and A_2 the area of the output end. Then

$$dQ = B(A_2/L^2)dA_1 \tag{3}$$

and

$$Q = (BA_2/L^2)\int_{\text{source}} dA_1 = (A_1A_2/L^2)B. \tag{4}$$

It does not matter whether the source plane is at the large end of the tube or the small end because the light flux varies as the product of these areas. This result also applies to the case of a uniform tube which is fitted with aperture diaphragms at the ends.

UNIFORM TUBE WITH ONE LENS AT ITS FOCAL DISTANCE FROM THE SOURCE PLANE

In this case the light which originates on a small element dA of the source and which falls on the

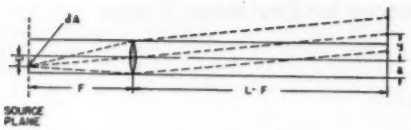

FIG. 1. Geometry for case of one lens placed at its focal length from the source.

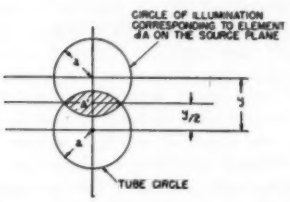

FIG. 2. Geometry at output end of tube.

lens is rendered into a parallel beam of light with a circular cross section of radius "a" as shown in Figs. 1 and 2. Since there is no reflection from the tube walls the amount of light from the element dA which gets through the output end of the tube is

$$dQ = BdA(A'/F^2) \tag{5}$$

where F is the focal length of the lens and A' is the area common to the tube circle at the output end of the tube and the circle of illumination as shown in Fig. 2. Hence the total light flux is

$$Q = B/F^2 \int_{y=0}^{y=y_2} A'(y)dA(y) \tag{6}$$

where A' and dA are both functions of y, and y_2 is the value of y at which the area common to both circles becomes zero (i.e., $y_2 = 2a$), or the value at which y', as shown in Fig. 1, becomes "a," whichever occurs first. Introducing the dimensionless parameter, $z = y/2a$, the area A' may be expressed as,

$$A'(z) = 2a^2[\cos^{-1}z - z(1-z^2)^{\frac{1}{2}}], \tag{7}$$

and the annular element of area $dA(z)$ on the source plane is,

$$dA(z) = 8\pi a^2 p^2 z dz, \tag{8}$$

where p is defined as

$$p = F/(L-F). \tag{9}$$

Now the integral given in Eq. (6) becomes,

$$Q = B/F^2 \int_0^{z_2} 16\pi a^4 p^2[\cos^{-1}z - z(1-z^2)^{\frac{1}{2}}]z dz$$

$$= (16\pi a^4 B/L^2)(1+p)^2 \times \int_0^{z_2} z[\cos^{-1}z - z(1-z^2)^{\frac{1}{2}}]dz. \tag{10}$$

If $p < \frac{1}{2}$, $z_2 = 1$, and if $p > \frac{1}{2}$, $z_2 = \frac{1}{2}p$. For $p < \frac{1}{2}$,

$$Q = (\pi^2 a^4 B/L^2)(1+p)^2 = Q_0(1+p)^2 \tag{11}$$

and for $p > \frac{1}{2}$

$$Q/Q_0 = \left(\frac{1+p}{p}\right)^2 \left\{1 - \frac{1}{\pi}\left(\frac{1+2p^2}{2p}\right)\left(1-\frac{1}{4p^2}\right)^{\frac{1}{2}} + \frac{2}{\pi}(p^2-1)\sin^{-1}\frac{1}{2p}\right\}. \tag{12}$$

As p goes to infinity (i.e., as the lens approaches the output end of the tube) the expression for Q in Eq. (12) is approximated very closely by,

$$Q/Q_0 = [1+(2/p)][1-(11/8\pi p)], \tag{13}$$

and thus it is seen from Eqs. (11) and (13) that if the lens is at the input or output ends of the tube the amount of light transmitted is the same as for the plane tube without lenses, while for intermediate positions the lens gives some increase. The values of Q/Q_0 for several values of p and F/L are plotted in Fig. 3. From this it is seen that the amount of transmitted light reaches a rather sharp maximum when the lens is placed at a little over 40 percent of the way from the input end to the output end of the tube. This arrangement was tested experimentally in the laboratory. The results are indicated by the x's on Fig. 3.

In the practical use of periscope systems for studying the light emitted from different parts of a flame structure it is important to know the contribution of each element of the flame structure to the total light transmitted by the periscope. For this reason an analysis of the contributions of ring-shaped elements of different radii on the light source plane has been made.

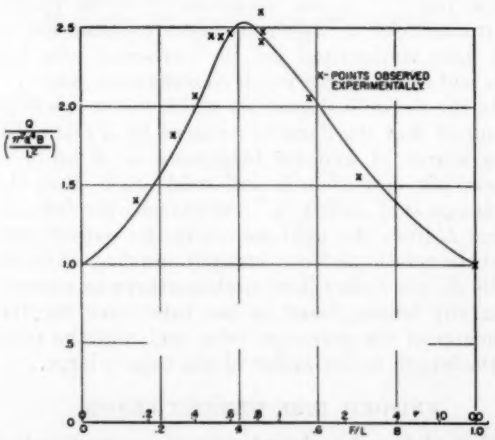

FIG. 3. Effect on transmission of light due to one lens placed at the focal length from the source.

In the plane of the light source let $z'=y'/a$. The relationship between z' and the z used above is

$$z'=2pz. \tag{14}$$

The contribution of the ring element $2\pi z'dz'$ to the light transmitted by the periscope tube is

$$dQ'=Q_0\cdot dz'(4Uz'/\pi)(1+p/p)^2, \tag{15}$$

where

$$U=\cos^{-1}\cdot z-z(1-z^2)^{\frac{1}{2}}$$

$$=\cos^{-1}\frac{z'}{2p}-\frac{z'}{2p}\left(1-\frac{z'^2}{4p^2}\right)^{\frac{1}{2}}. \tag{16}$$

The graphs for $dQ'/(Q_0dz')$ for different values of p are plotted against z' in Fig. 4. From this it is seen that when the lens is near the input end of the tube most of the transmitted light originates near the optic axis whereas when the lens is near the output end most of it originates near the periphery of the tube circle. For the position of maximum transmission, *viz.*, for $p=0.7$ or $F/L=0.41$, nearly half of the transmitted light originates within a circle of half the radius of the tube. This relationship was tested experimentally for the cases $p=0.2$, $p=0.5$, and $p=1.0$ by using the same apparatus mentioned above and placing circular apertures of various diameters on the source plane. Since this gave the total light transmitted from a circular aperture it was necessary to graphically integrate the curves given in Fig. 4 to find the theoretical amount of transmitted light in each case. The integral curves along with the experimentally observed points are shown in Fig. 5. The lack of agreement between the theoretical curves and the experimental points is within the experimental error.

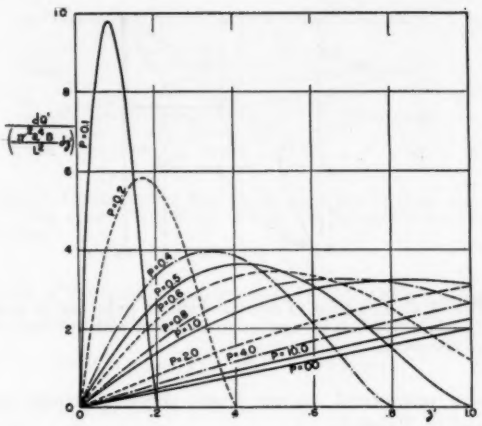

FIG. 4. Contribution of ring elements $z'dz'$ of the light source to the light transmitted by a periscope tube with one lens placed at its focal length from the light source.

ONE LENS OF $F=L/4$ PLACED AT $L/2$

In this case the light source is at a distance $2F$ from the lens and consequently a real image is formed at the output end of the tube. The ray diagram presented in Fig. 6 shows that all of the light which originates on a small element of the source and which strikes the lens is transmitted to the output end of the tube. Hence each element of area dA on the source plane contributes an amount

$$dQ=B[\pi a^2/(L/2)^2]dA \tag{17}$$

to the light transmitted by the tube. Hence the total amount transmitted is

$$Q=(4\pi^2a^4B/L^2)=4Q_0 \tag{18}$$

where Q_0 is the amount transmitted without any lens. Since each element of area on the source plane contributes the same amount, most of the total transmitted light originates near the periphery of the tube circle. This is not desirable in those cases in which the light should be drawn from near the optic axis.

UNIFORM TUBE WITH LENSES SPACED $2F$ APART

Consider first the arrangement illustrated in Fig. 7 in which the first lens of the series is placed at a distance of $2F$ from the source plane. A full-sized image of the source then falls on lenses 2, 4, $6,\cdots 2n$. All of the light that falls on the first lens is transmitted to the output end of the tube except for that which is lost in the lenses by reflection and absorption.

The amount of light that falls on the first lens is

$$Q_1=(\pi^2a^4B/4F^2)=[\pi^2a^4B/(L/n)^2], \tag{19}$$

where n is the number of lenses used (i.e., the number of intervals into which the tube is divided by the lenses). If each lens transmits a fraction x of the light incident upon it, then the amount transmitted by n lenses in series is x^n times that incident upon the first lens. Hence, in a system of the type shown in Fig. 7 the total transmitted light is

$$Q=(\pi^2a^4B/L^2)n^2x^n. \tag{20}$$

Since x is always less than unity the factor n^2x^n does not increase indefinitely but reaches a definite maximum for any given value of x. By differentiation it is found that the n which corresponds to the maximum is given by

$$n_{\max}=-2/\ln x, \tag{21}$$

and the maximum value of n^2x^n is

$$(n^2x^n)_{\max}=(2/\ln x)^2e^{-2}. \tag{22}$$

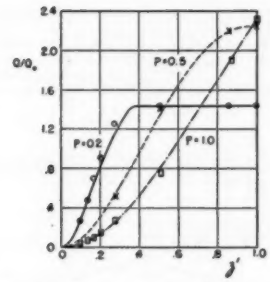

FIG. 5. Contribution of circular zones of radius ax' in terms of $Q_0=(\pi^2a^4B/L^2)$. This applies to a system with one lens at F from the source.

Some numerical values from these relations are listed in Table I.

In practice it is not difficult to get lenses which transmit 90 percent of the light incident upon them and hence it is possible by using 19 lenses to increase the light transmission of a given tube by a factor of about 50. In many cases, however, it is not convenient or practical to mount so many lenses in the tube and under such circumstances a compromise must be made between a reasonable number of lenses and the amount of transmitted light.

Also in practical applications it is not always convenient to use a system which can be adjusted to make the distance between the light source and the first lens exactly $2F$. It is of interest, therefore, to find how the total amount of transmitted light depends upon the distance between the light source and the first lens.

The case which is most easily analyzed is that in which the source is at a distance F from the first lens. As shown in Fig. 8, all the light that strikes the second lens will continue through to the end of the tube. Thus the analysis is exactly the same as that given in the preceding section for the single lens at F from the source, with $p=\frac{1}{2}$ and $L=3F$. By substituting these values in Eq. (11) one gets,

$$Q=[\pi^2a^4B/(3F)^2](1+\tfrac{1}{2})^2=(\pi^2a^4B/4F^2, \quad (23)$$

which, as will be noted by comparison with Eq. (19), is the same as for the case in which the source is at a distance $2F$ from the first lens. To get the actual amount of light transmitted the Q given in Eq. (23) must be multiplied by the lens transmission factor given in Eq. (20).

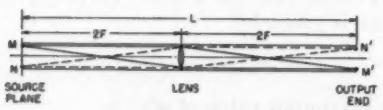

FIG. 6. System with single lens of $F=L/4$ placed at the midpoint of the tube.

TABLE I. Values of n_{max} and $(\pi^2x^n)_{max}$ for various values of x.

x	$n_{max}=-2/\ln x$	$(\pi^2x^n)_{max}$
1.00	∞	∞
0.95	39.0	206
0.90	19.2	49.6
0.85	12.3	20.6
0.80	9.9	13.4
0.75	7.0	7.8
0.70	5.7	4.3

The general case in which the source plane is at any distance from the first lens is more difficult to analyze. The geometry used in the analysis is presented in Fig. 9. Consider the element M on the optic axis at the source plane. Of the rays that originate in M only those which fall within the cone $MM_1'M_2'$ fall eventually on the second lens and hence are transmitted. This cone is determined by the limiting rays M_2M and M_1M which may be considered as coming from a virtual source plane placed $2F$ from the first lens. Any ray which emerges from the virtual point M_2 and which passes through the first lens comes to the point M_2'' on the second lens, and likewise for M_1 and M_1''.

The same criterion may be applied to the source element N. As indicated in Fig. 9 the base of the cone of transmission for this element comes tangent to the rim of the first lens. Hence for any element, 0, which falls farther from the optic axis than N the base of the cone of transmission $0_1'0_2'$ falls partially outside the first lens and hence only a fraction of the rays which fall within the cone are transmitted. The analysis may be made on the basis of the solid angle subtended at the element of area on the source plane by the area common to the first lens circle and the circular base of the cone of transmission.

For convenience the following dimensionless parameters will be used:

$$\alpha=x/2F,$$
$$\beta=(2F-x)/2F, \quad (24)$$
$$\gamma=x/(2F-x).$$

The base of the cone of transmission which lies in the plane of the first lens will be referred to as the circle of illumination. The radius of this circle is γa and the solid angle it subtends at the source plane is $\pi\gamma^2a^2/x^2$. When the circle of illumination falls partially outside the tube circle the effective

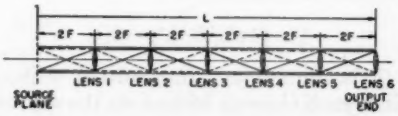

FIG. 7. System with lenses spaced $2F$ apart and with the first lens $2F$ from the source plane.

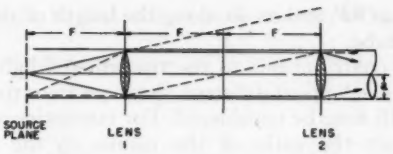

FIG. 8. Input end of a "2F" system with the source at a distance of F from the first lens.

solid angle is the ratio of the area common to both circles to πa^2.

The general case for the geometry of these circles is illustrated in Fig. 10. The distance between centers of the tube circle and the circle of illumination in the plane of the first lens is denoted by y. It is found that the areas A_1 and A_2 (Fig. 10) are given by the expressions:

$$A_1=a^2\cos^{-1}\left[\frac{y^2+a^2(1-\gamma^2)}{2ay}\right]-\frac{y^2+a^2(1-\gamma^2)}{2y}\left\{a^2-\left[\frac{y^2+a^2(1-\gamma^2)}{2y}\right]^2\right\}^{\frac{1}{2}}$$

$$A_2=\gamma^2a^2\cos^{-1}\left[\frac{y^2-a^2(1-\gamma^2)}{2\gamma ay}\right]-\frac{y^2-a^2(1-\gamma^2)}{2y}\left\{\gamma^2a^2-\left[\frac{y^2-a^2(1-\gamma^2)}{2y}\right]^2\right\}^{\frac{1}{2}} \quad (25)$$

and hence the total area common to the two circles is:

$$A=A_1+A_2=a^2\left\{\cos^{-1}\left[\frac{y^2+a^2(1-\gamma^2)}{2ay}\right]+\gamma^2\cos^{-1}\left[\frac{y^2-a^2(1-\gamma^2)}{2\gamma ay}\right]\right\}-\tfrac{1}{2}\{-a^4(1-\gamma^2)^2+2a^2(1+\gamma^2)y^2-y^4\}^{\frac{1}{2}}. \quad (26)$$

The expression for the total transmitted light is now found by integrating over all values of y up to the value for which the circle of illumination falls outside the tube circle. The circle of illumination falls entirely within the tube circle for $y=0$ to $y=a(1-\gamma)$, and partially within for $y=a(1-\gamma)$ to $y=a(1+\gamma)$. The radius y in the plane of the first lens corresponds to a radius of $y'=\beta y$ in the source plane. Thus the annular element of area on the source plane which corresponds to a given y is $2\pi y'dy'=2\pi\beta^2ydy$. Hence,

$$Q=\frac{\pi\gamma^2a^2B}{x^2}\cdot 2\pi\beta^2\int_0^{a(1-\gamma)}ydy+B2\pi\beta^2\int_{a(1-\gamma)}^{a(1+\gamma)}\frac{A}{x^2}ydy, \quad (27)$$

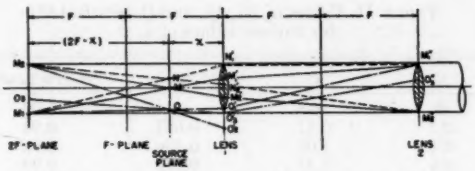

FIG. 9. Geometry for analysis of entrance of light into a "2F" lens system where the source is a distance x from the first lens.

where A is the expression given in Eq. (26). Making the substitutions

$$x=2F(\gamma/\gamma+1) \quad \text{and} \quad \beta=(1/\gamma+1), \quad (28)$$

and integrating, one gets,

$$Q=\frac{\pi^2a^4B}{4F^2}\left[\gamma^2+\frac{2}{\gamma}-\frac{2}{\pi\gamma^2}V_1-\frac{2}{\pi}V_2\right], \quad (29)$$

where,

$$V_1=\int_{1-\gamma}^{1+\gamma}z\sin^{-1}\left(\frac{z^2+1-\gamma^2}{2z}\right)dz,$$

$$V_2=\int_{1-\gamma}^{1+\gamma}z\sin^{-1}\left(\frac{z^2-1+\gamma^2}{2\gamma z}\right)dz. \quad (30)$$

$$z=y/a.$$

The integrals V_1 and V_2 are not integrable by analytical methods (except for the case $\gamma=1$) and consequently it has been necessary to integrate them graphically. The range over which the integrals are real is $\gamma=0$ to $\gamma=1$. The other range of interest in this case, $\gamma=1$ to $\gamma=\infty$, will be discussed later. The values of the integrals V_1 and V_2 and the ratio $Q/(\pi^2a^4B/4F^2)$ are given in Table II. The value of the bracket in Eq. (29) is determined by a relatively small difference between the large second and third terms, and hence the values from the numerical integration which are listed in Table II may be in error by a few percent. Evidence will be presented later which indicates that the true value is constant and equal to unity for all values of γ.

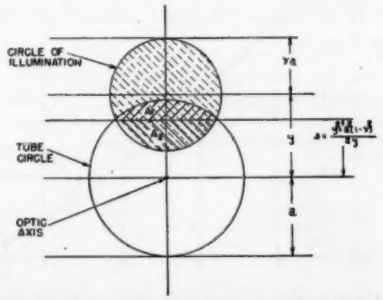

FIG. 10. Geometry of tube circle and circle of illumination.

TABLE II. Values of V_1, V_2, and $Q/(\pi^2a^4B/4F^2)$ for various values of γ.

γ	V_1	V_2	$Q/(\pi^2a^4B/4F^2)$
0	0	0	1.00
0.2	0.57	0.067	0.94
0.4	1.01	0.256	0.98
0.6	1.33	0.572	0.99
0.8	1.52	1.03	0.97
1.0	$\pi/2$	$\pi/2$	1.00

The values for $\gamma=0$ and $\gamma=1$ come from exact integrations.

It will now be proved that the light transmission characteristics are identical for the two positions in which the source plane is at a distance x or $2F-x$ from the first lens. This means that the transmission characteristics are symmetrical about the F-plane and whatever is true for the position γ is also true for the position $\gamma'=1/\gamma$.

To see this refer again to Fig. 9. Consider the element N on the source plane. If there were a lens in the $2F$-plane the ray NM_2 would be bent back to the same angle to the optic axis as the ray $M_2''N_2'$. Likewise, the ray NM_1 which goes through the focus would be bent parallel to the optic axis the same as ray $M_1''N_1'$. Consequently, any ray from N which falls within the cone NM_1M_2 would be transmitted just the same as any ray falling in the cone $NN_1'N_2'$. The same considerations are true for the element O on the source plane where the points M_1 and O_3 correspond to the points M_1'' and O_3'', respectively. This means that if a source plane be placed at any arbitrary position in a "$2F$" lens system the amount of light transmitted from any given element of that source plane is the same to the left as to the right. Consequently, the total light flux is the same in both directions and the strength of the contributions of elements at given distances from the optic axis is the same. It is also evident that an inverted image of the source plane is formed at a distance $4F$ from it, an upright image at $8F$, and so on along the length of the periscope tube.

The contributions to the transmitted light from source elements at different distances from the optic axis will now be considered. For convenience let z' represent the ratio of the radius of the source element to the tube diameter. This parameter is related to the y coordinate in the plane of the first lens as follows:

$$z'=\beta(y/a)=[y/a(1+\gamma)]. \tag{31}$$

In the region $z'=0$ to $z'=(1-\gamma)/(1+\gamma)$ the circle of illumination falls entirely within the tube circle and hence the contribution of any element of area dA on the source plane is

$$dQ_1=(\pi a^2BdA/4F^2)(1+\gamma)^2. \tag{32}$$

In the region $z'=(1-\gamma)/(1+\gamma)$ to $z'=1$, the circle of illumination falls only partially within the tube circle and the contribution is,

$$dQ_2=(\pi a^2BdA/4F^2)(A/\pi a^2)(1+\gamma/\gamma)^2, \tag{33}$$

where A is the expression given in Eq. (26), which when put in terms of z' becomes:

$$A=\frac{\pi a^2}{2}\left\{(1+\gamma^2)-\left(\frac{1+\gamma}{\pi}\right)[-(1-\gamma)^2+2(1+\gamma^2)z'^2 -(1+\gamma)^2z'^4]^{\frac{1}{2}}-\frac{2}{\pi}\sin^{-1}\left[\frac{(1+\gamma)z'^2+1-\gamma}{2z'}\right] -\frac{2\gamma^2}{\pi}\sin^{-1}\left[\frac{(1+\gamma)z'^2-(1-\gamma)}{2\gamma z'}\right]\right\}. \tag{34}$$

The values of the variables $(1+\gamma)^2$ and $(A/\pi a^2)(1+\gamma/\gamma)^2$ which appear in Eqs. (32) and (33) are plotted for various values of γ in Fig. 11. From this

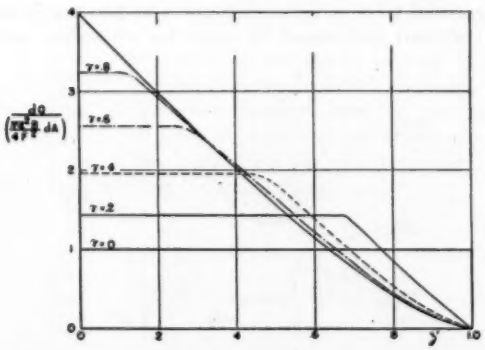

FIG. 11. Contribution to transmitted light from elements of equal area on the source plane.

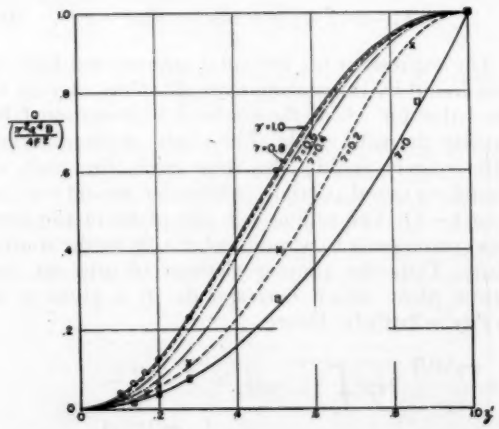

FIG. 12. Contribution to total transmitted light due to circular zones of radius z' on source plane.

TABLE III. Values of the ratio $Q/(\pi^2a^4B/4F^2)$ for various values of γ.

γ	$Q/(\pi^2a^4B/4F^2)$
0	1.00
0.2	1.01
0.4	1.01
0.6	1.01
0.8	1.00
1.0	1.00

it is seen that for the case $\gamma=1$ (i.e., when the source plane is one focal distance from the first lens) the center elements of the source plane contribute much more to the transmitted light than do the elements near the periphery, and that for the case $\gamma=0$ (i.e., when the source plane is $2F$ from the first lens) all the elements of the source plane contribute equally.

The contributions of circular zones of fractional radius s' may be found by multiplying the expressions in Eqs. (32) and (33) by $2a^2s'ds'$ and integrating from 0 to s'. This integration has been done graphically with the results shown in Fig. 12. It should be noted that for all the values of γ for which the operations were carried out the value of the ratio $Q/(\pi^2a^4B/4F^2)$ approaches very close to unity at $s'=1$, that is, for full aperture. The actual values obtained are listed in Table III. These values, together with those given in Table II, show quite definitely that the total amount of light transmitted is constant for any position of the source plane between 0 and $2F$ of the first lens. This conclusion is verified by experimental measurements. The experimental values for the amount of light transmitted from circular apertures of various values of s' for the cases $\gamma=0$, $\gamma=0.2$ and $\gamma=1.0$ are indicated in Fig. 12. The agreement between the experimental points and the theoretical curves is within the experimental error in all cases.

TUBES WITH REFLECTING SIDE WALLS

In all the transmission systems described above it has been assumed that any light which strikes the side walls is absorbed. The case in which the side walls are reflecting will now be analyzed. The geometry is indicated in Fig. 13. If a ray of light starts from the source at an angle θ to the optic axis the distance between reflections is $d=2a\ \mathrm{ctn}\theta$ and the number of reflections along the tube is

$$n=L/d=(L/2a)\tan\theta. \tag{35}$$

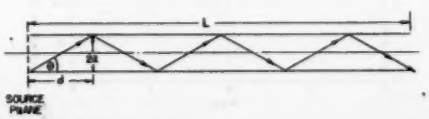

FIG. 13. Reflection of light along walls of periscope tube.

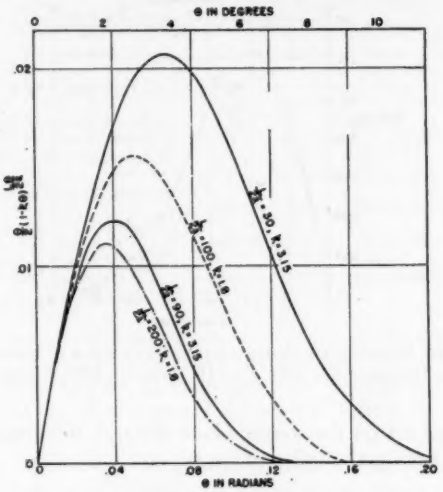

FIG. 14. Angular distribution of light transmitted by a tube with reflecting walls.

Let $x(\theta)$ represent the fraction of light which is reflected at each reflection of the ray of light from the wall. Then the fraction of the original ray which is transmitted is $[x(\theta)]^n$. From data reported in the literature it is found that for values of θ less than about 15° the fraction of light reflected at each incidence is given closely by the linear relationship,

$$x(\theta)=1-k\theta. \tag{36}$$

When θ is expressed in radians the values of k for sodium yellow light reflected from polished steel, copper, silver (or aluminum), and glass are, respectively, 1.8, 1.4, 0.36, and 3.15. Using the relationships given in Eqs. (35) and (36), and expressing the small angles in radians, the fraction of the original ray which is transmitted is

$$T(\theta)=(1-k\theta)^{(L/2a)\theta}. \tag{37}$$

Again let the brightness of the source plane be B units of power per unit of area and solid angle. The amount of light which passes from the element

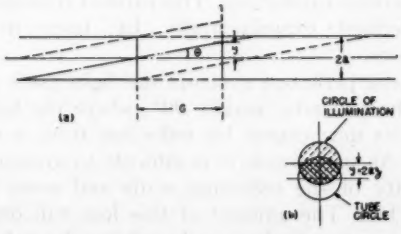

FIG. 15. Ray diagram for loss of light at gap in reflecting walls of tube.

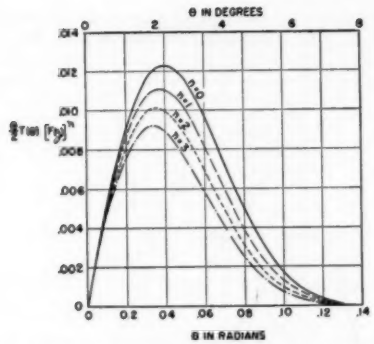

FIG. 16. Values of transmitted light as a function of the angle for the case $L/2a=90$, $k=3.15$, $G/2a=2$, with n gaps.

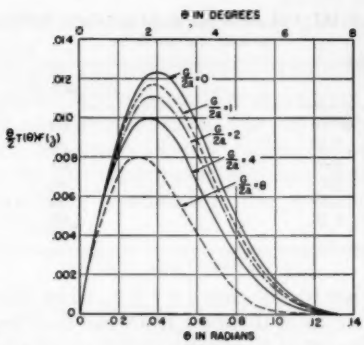

FIG. 17. Values of transmitted light for a glass tube for which $L/2a=90$, $k=3.15$, and with one gap of various lengths.

of area dA on the source plane through the annular zone $\theta d\theta$ on the unit sphere is

$$dQ=2\pi B\sin\theta d\theta\approx 2\pi B\theta d\theta dA, \tag{38}$$

and the amount which is transmitted is

$$dQ'=dA\,2\pi B\theta(1-k\theta)^{(L/2a)\theta}d\theta. \tag{39}$$

The total amount of light transmitted is now found by integrating Eq. (39) over the circle of radius a on the source plane and over all angles θ which are of importance,

$$Q=2\pi B\int_{\substack{\text{source}\\\text{plane}}}\int_{\substack{\text{reflection}\\\text{angles}}}\theta(1-k\theta)^{(L/2a)\theta}d\theta dA$$

$$=2\pi^2a^2B\int\theta(1-k\theta)^{(L/2a)\theta}d\theta. \tag{40}$$

The values of the last integrand for several values of $L/2a$ are plotted in Fig. 14. The most striking characteristic shown by these curves is the concentration of the light at relatively samll angles to the optic axis. For example, for the case of $L/2a=90$, $k=3.15$, over 90 percent of the transmitted light falls within five degrees of the optic axis and theoretically the amount of light transmitted is about a hundred times as great as if the walls were non-reflecting. This ratio of transmission was verified experimentally by tests in the laboratory.

In some periscope systems the light path must make sharp bends, usually 90°, where the light is set on its new course by reflection from a plane mirror. At such bends it is difficult to arrange for continuity of the reflecting walls and some light will be lost. The amount of this loss will depend upon the ratio of the gap length to the tube diameter, and the angle of the light rays.

The geometry used in determining the fraction of the light transmitted past a gap is shown in Fig. 15. The gap has a length G. The parameter $z=y/2a\approx G\theta/2a$ is introduced for convenience. It is found that the fraction of the light transmitted at the angle corresponding to z is:

$$F(z)=(2/\pi)[\cos^{-1}z-z(1-z^2)^{\frac{1}{2}}]. \tag{41}$$

For values of z less than 0.2 this is approximated closely by the series,

$$F(z)=1-(4/\pi)z+z^2/\pi. \tag{42}$$

The amount of the gap loss does not depend upon the position of the gap along the length of the tube because the coefficient of reflection is independent of the intensity of the light. Hence the amount of light transmitted at any angle θ when there are n equal gaps in the tube is

$$dQ'=dA\,2\pi B\theta\cdot T(\theta)\cdot[F(z)]^n d\theta. \tag{43}$$

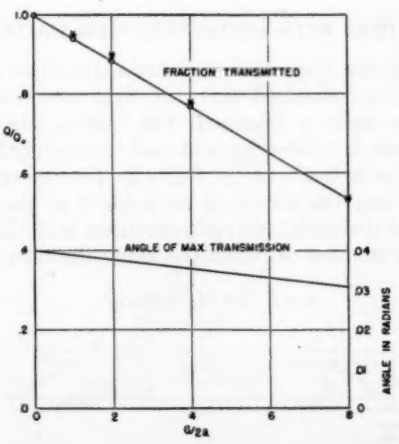

FIG. 18. Effects of single gap (derived from Fig. 17).

Some values for the special case of $L/2a=90$, $k=3.15$, $G/2a=2$ are presented in Fig. 16. The effect of the gaps is to decrease the total amount of transmitted light at all angles, and to decrease the angle of maximum transmission. This is a desirable effect if a close approximation to parallel light is desired at the output end.

The effect of a single gap of various lengths is shown in Figs. 17 and 18. The loss of light caused by the presence of a gap is not so great as one might expect. This is due to the fact that a large fraction of the light is transmitted at quite small angles to the optic axis. Quantitative measurements were made of the light transmitted by a glass tube with an $L/2a$ ratio of 90 for gaps of 0, 1, 2, 4 and 8 diameters. The results of these measurements are shown by asterisks in Fig. 18. In this figure Q/Q_0 indicates the ratio of the total light transmitted with the gap to that without the gap. The agreement between the experimental points and the theoretical curves shows that the theory is reasonably correct.

SUMMARY AND CONCLUSIONS

The maximum amount of light which can be transmitted (in terms of that transmitted by a plain black-walled tube), and the type of image formed by the systems which have been described in this report, are listed in Table IV. The results of the theoretical analyses and the experiments described in this report lead to the following conclusions:

TABLE IV. Light transmitted by various systems.

Arrangement	Light transmitted	Image of source
Uniform tube, non-reflecting walls	Q_0	None
One lens, $F=0.42L$, placed at $0.42L$ from source plane	$2.5\ Q_0$	None
One lens, $F=0.25L$, placed at $0.5L$	$4.0\ Q_0$	Sharp, normal size
Two lenses, $F=L/6$, placed at $L/3$ and $2L/3$	Approx. $7.3\ Q_0$	Sharp, normal size
Three lenses, $F=L/8$, placed at $L/4$, $2L/4$, and $3L/4$	Approx. $11.7\ Q_0$	Sharp, normal size
Four lenses, $F=L/10$, placed at $L/5$, $2L/5$, $3L/5$, and $4L/5$	Approx. $16.4\ Q_0$	Sharp, normal size
No lenses, glass reflecting walls, $L/2a=90$	Approx. $100\ Q_0$	None

(1) The most direct way to increase the amount of transmitted light is to increase the diameter of the tube to a maximum practical value while its length is kept to a minimum.

(2) A tube with reflecting side walls give s the maximum possible amount of transmitted light, but is satisfactory only in those cases in which it is not necessary to form an image of the source.

(3) In case it is necessary to form a well-defined image of the source at the output end of the tube, the "$2F$" type of lens system is best.

Infra-Red Absorption Analyses Used in Control of a Pilot-Plant Process

D. L. FRY
Research Laboratories Division, General Motors Corporation, Detroit, Michigan
(Received October 11, 1948)

The methods described herein are of general interest because they illustrate the effectiveness of infra-red in the analysis of hydrocarbons and the control of experimental and pilot-plant processes. Some particular problems are discussed. These problems vary in complexity from analyzing for two components in a mixture of two, to analyzing for three components in a mixture of twelve or six in a mixture of six. In some cases the analyses are possible only for certain ranges of concentration; in other cases the concentrations of the components can vary from 0 to 100 percent.

The analyses were determined with an accuracy of plus or minus 1 percent, providing the range of concentration of the various components stayed within the limits for which the method was calibrated.

In some cases it was necessary to set up calibration curves with inadequate standard samples. It is most interesting that even under these adverse conditions it was possible to attain reasonable accuracy and to make a contribution to the problem of controlling the pilot-plant. In control problems it is frequently not important to know the absolute composition; it is sufficient to know relative values of concentration from day to day.

INTRODUCTION

IN the process of manufacturing 2,2,3-trimethylbutane (triptane),[1] analyses were required quickly and in large numbers. The conventional methods, which depended upon distillation, adsorption, index of refraction, aniline point, and freezing point, were too slow or not accurate enough. It was therefore necessary to make use of some other method. The infra-red absorption method was selected. The work described in this paper was completed in 1945. The instrumentation used has been previously described.[2] All the quantitative work done made use of the "cell-in cell-out" techniques as described with extensive references in a recent article.[3]

There are about 20 different processes in the manufacture of triptane and 15 of them were controlled to some extent by infra-red. A list of the compounds encountered in the course of this work and the problem number in which they were involved is included in Table I.

As space does not permit a description of the total problem, this paper will be limited to a discussion of some of the variables which control the accuracy attainable. This will not include a discussion of instrumental errors; it has been found that errors in results directly attributable to absorption measurements seldom exceed ±0.5 percent in the final analysis.

DISCUSSION

Calibration with Impure Compounds

Usually, the first step in setting up an infra-red procedure is to obtain pure components. Of the 30 or more components needed, only 12 could be obtained in pure form, 13 were obtainable in commercial or 90 percent purity grades, and the rest were unattainable. The use of the pure compounds is obviously desirable because it simplifies calibration. However, considerable work can be done with the less pure standards as will be illustrated in problem 1 (see Table I).

2-methyl-2-butanol for use as a standard was not available, so the infra-red spectrum (1–15μ) of an impure sample was obtained. A strong band at 9.6μ was chosen as the analytical position. The absorption at this position of several samples was obtained. The sample absorbing the most was assumed the most pure. Spectra of the least pure and the most pure were obtained and the differences noted. A strong band at 10.9μ present in the least pure, but not in the most pure, was selected as a band due to impurity. From consideration of the boiling points and the methods of manufacturing, it was decided that the two most likely impurities were 2 chloro-2-methylbutane and 2-butanol. These two compounds were obtained in an impure condition and their spectra recorded. It was immediately obvious that the main impurity in the 2-methyl-2-butanol samples was 2-butanol.

A mixture of 25 percent 2-butanol (impure) and 75 percent 2-methyl-2-butanol (most pure) was made, and the absorptions at 9.6 and 10.9μ of this mixture and of the two unblended samples were obtained. These data were used to plot two calibration curves. (If no other compound were present, only one curve would be necessary; the other component could be found by differences.) The calibration curves were plotted as if the alcohols were 100 percent pure. Actually the true values in the standards were less than 100 percent; therefore, the true values of the alcohols in unknowns were

[1] Verle A. Miller and W. G. Lovell, Ind. Eng. Chem. **40**, 1138 (1948).
[2] D. L. Fry, R. E. Nusbaum, and H. M. Randall, J. App. Phys. **17**, 150 (1946).
[3] Van Zandt Williams, Rev. Sci. Inst. **19**, 135 (1948), see especially p. 145.

less than the values read from the calibration curves.

Even though highly accurate analyses of this type were not possible, the results were correlated with reaction efficiency with resulting improvement thereof.

Overlapping Bands

Examination of Table I shows that in problem 5, a mixture of 12 compounds was analyzed for 3. A comparison of the spectra of the compounds involved in this mixture showed that overlapping of bands was so serious that suitable wave-length positions could not be found for the analysis of the three compounds. However, it was found that this situation could be corrected by removing the low boiling materials. It was a simple matter for the pilot-plant chemists to "degas" the sample. The absorption measurements were then made on the degassed sample.

An average sample of problem 5 is shown in Table II; the average portion discarded was 10 percent.

The percent concentrations of the 3 components reported were the concentrations in the total sample before degassing. Therefore, if the amount of the material discarded from unknowns in the degassing process varied from 10 percent by ±6 percent of the total sample, an error of ±1 percent was introduced into the amount of 2,3,3-trimethylbutene reported, and similar errors were caused in the analyses for the other compounds.

Although the degassing consumed an undesirable amount of time and introduced a source of error, it made possible an analysis which otherwise could not be obtained.

Unseparated Interfering Components

In addition to the errors connected with the degassing process in problem 5, there were errors which resulted from variations in the concentrations of the components in the degassed portion of the mixture. For the first part of this discussion we will assume that 2,3,3-trimethylbutene is the only compound analyzed for and that readings are made at only one wave-length. The absorption of the degassed portion of an unknown can be considered as the sum of the absorptions of parts 2 and 3 of Table II. In order that the analysis be accurate, the absorption of part 2 at the selected wave-length must be essentially constant from sample to sample while the absorption due to 2,3,3-trimethylbutene changes with the concentration. If a 1 percent deviation in the analysis for 2,3,3-trimethylbutene is tolerated, the absorption of part 2 can vary by the amount of absorption equal to that of 1 percent 2,3,3-trimethylbutene.

TABLE I. Compounds involved in the pilot-plant process.

Compounds	Boiling points °C	Problems in which they are involved*
2-methyl-2-butanol	102	*1* *2*
2-butanol	100	*1* *2*
2-chloro-2-methylbutane	86	1 2
1-butene	2	3 4
2-butene	2	3 4
3-methyl-1-butene	20	*2* 3 4 5 6
2-methyl-1-butene	31	*2* *3* *4* 5 6
2-methyl-2-butene	39	*2* *3* *4* *5* *6* *11*
2,3-dimethyl-1-butene	56	*4* 5 6 *7* *8* *11*
2,3-dimethyl-2-butene	73	*5* *6* *7* *8* *9* *10* 11 13
2,3,3-trimethylbutene	78	*5* *6* 7 8 *9* *10* 11 *12*
3-methyl-2-pentene (LB)	68	5 6 7 8 9 10 11
3-methyl-2-pentene (HB)	71	5 6 7 8 9 10 11
2-ethyl-1-butene	65	5 6 7 8 9 10 11
Hexenes**	50–70	5 6 7 8 9 10 11
Heptenes***	60–90	5 6 12
Octene cut****	Above 80	5 6 12
2,3-dimethylbutane	58	*13*
2-methylpentane	60	13
3-methylpentane	63	13
2,3-dimethylpentane	90	*14*
2-methylhexane	90	14
3-methylhexane	89	14
3,3-dimethylpentane	86	14
2,2,3-trimethylpentane	110	*15*
2,3,4-trimethylpentane	114	15
2,3,3-trimethylpentane	115	15
2,2,4-trimethylpentane	113	15

* Italicized numbers mean the sample is analyzed for this compound.
** Exactly which hexenes are present is not known. They make up less than 2 percent of any of the samples.
*** Exactly which ones are present is not known. They make up less than 5 percent of a type 5 sample but may be as high as 50 percent of type 12 sample.
**** This probably is composed of more than octenes. It makes up less than 5 percent of a type 5 sample but may be as high as 50 percent of a type 12 sample.

This effect is further described in Table III. Any single variation shown in column 2 can cause a ±1 percent error in the 2,3,3-trimethylbutene analysis. All percentages refer to percent of total sample. In some cases, obviously, enough negative variation to cause a 1 percent error cannot occur because there isn't enough of the material present in the average sample.

Following this reasoning for all the compounds at the three wave-lengths involved permits a determination of the maximum amount of variation allowable for each compound in order that the error in the determination of any of the three is less than 1 percent. This information is shown in Table IV. If the allowable error were 2 percent, every number in the table would double. Data shown in Table IV can be obtained from the optical density of the components at the desired wave-lengths.

There are other similar sources of errors which are of an indeterminate amount. For instance, octene cut is treated as a pure material. Actually it is a mixture and may vary in itself from one sample to another. Also, if compounds which are not listed are present, they may have strong absorptions at the three wave-length positions and may therefore cause errors.

Calibration with Analyzed Samples

In problem 5 a difficult situation was also encountered with regard to standards for calibration

TABLE II. Composition of a typical sample of mixture encountered in problem 5.*

	Compound	Percent of total sample
(1) Portion discarded (10%)	Methyl chloride	2.7
	Methyl ether	5.0
	Amylenes	2.3
(2) Remaining part (78%)	Amylenes	40.4
	2,3-dimethyl-1-butene	16.1
	2,3-dimethyl-2-butene / Methylpentenes	12.7
	Octene cut	8.8
(3) 2,3,3-Trimethylbutene (12%)	2,3,3-trimethylbutene	12.0
Total 100%		100.0%

* Parts (2) plus (3) make up the degassed portion.

curves. Some of the pure compounds involved in the analysis were available. However, if a synthetic mixture of only these compounds had been used for the calibration curves, no account would have been taken of the effects of the other constituents present. The proportion of these other constituents was not known but seemed to vary as a function of the amount of 2,3,3-trimethylbutene formed in the reaction.

In the general case, a problem in which there were so many uncontrolled variables would hardly seem solvable. However, the problem was actually simplified by the fact that when slight variations in the process occurred so as to change the amount of one constituent, the amounts of other constituents tended to change in some systematic related fashion.

It was therefore decided to use samples from the pilot-plant and to accept estimates from distillation experiments of the amount of the three constituents for which analysis was to be made. It was then assumed that the effects of the other constituents on the absorption at the chosen wave-lengths would be the same in the unknown samples as in the standards. Also, the wave-lengths were so chosen that only a small effect might be expected from the absorption of these other constituents.

Actually, the estimation of concentrations from distillation data, mentioned above, is not simple. The boiling points of 2-methyl-2-butene and 2-methyl-1-butene are close. These compounds were not completely separable in the stills and the analysis of the individual compounds is indeterminate from the distillation curves for small amounts of the materials. However, when the fractions containing the two amylenes from a large number of samples were combined and redistilled, an analysis of 66 percent of 2-methyl-2-butene was found. It was then assumed that in the *standards* there was the same ratio of 2-methyl-2-butene to 2-methyl-1-butene as in the average.

A similar difficulty was encountered in the distillation analyses of 2,3-dimethyl-2-butene in the presence of the three hexenes with close boiling points. An estimate of 60 percent 2,3-dimethyl-2-butene in the hexene fraction was taken for the standard samples.

The calibration curves were then obtained from the optical density measurements and the estimated distillation analyses.

In the case of the calibration curve for 2,3-dimethyl-2-butene, the points did not determine the slope because the variation in the concentrations was not great enough. However, in the narrow range of concentration of the unknowns, small variations in the slope will not cause much error. Therefore, the slope was approximately determined from the known absorption of pure 2,3-dimethyl-2-butene.

In the course of the work for the pilot-plant, 13 distillation checks were made on analyses of the problem 5 type with an average variation of ±0.5 percent in the 2,3,3-trimethylbutene content. There has been no way of checking exactly the accuracy of the absolute analyses for the other two components. However, changes made in the process, based on these analyses, gave better results in the final product.

Problem 6 was very similar to 5 except that the concentrations of the compounds were quite different (Table I). Also as improvements were made in the process, the amount of unusable material (octene cut) was reduced to a value outside the range for which the method was calibrated. It was then necessary to make a distillation check and to shift the calibration curve.

Calibration with Pure Samples

In one case, problem 7, Table I, all six constituents of the mixture were available in pure form (not including heptenes which were present in very small quantities). In this case the method of applied corrections[2,4] was used.

TABLE III. Variation to cause 1 percent error in infra-red analysis for 2,3,3-trimethylbutene in problem 5.

Component	Variation
Amylenes: 2-methyl-1-butene	±30.0%
Amylenes: 2-methyl-2-butene	±10.0
2,3-dimethyl-1-butene	±30.0
2,3-dimethyl-2-butene	±14.0
Methylpentenes	± 7.0
Octene cut	± 4.0

[4] R. R. Brattain, R. S. Rasmussen, and A. M. Cravath, J. App. Phys. 14, 418 (1943).

TABLE IV. The effect of variation in composition on the accuracy of analyses for problem 5.

Component	Average concentration	Variations to cause 1% error in determination of ** component 8.3μ	8.6μ	12.5μ	Over-all limits
2-methyl-2-butene	28	±10	±37	**	±10
2,3-dimethyl-2-butene	8	±14	**	±100	±14
2,3,3-trimethylbutene	12	**	± 6	±100	± 6
Gas portion (discarded)	10	± 6	±10	± 3	± 3
3-methyl-1-butene	1	± 8	±10	± 17	± 8
2-methyl-1-butene	12	±30	±89	± 14	±14
2,3-dimethyl-1-butene	16	±30	± 8	±100	± 8
2-ethyl-1-butene	1	± 5	±48	± 8	± 5
3-methyl-2-pentene (LB)	1	± 7	±36	± 4	± 4
3-methyl-2-pentene (HB)	2	± 6	±28	± 8	± 6
Octene cut	9	± 4	± 4	± 13	± 4

** Mixture analyzed for this compound at this wave-length.

In testing the method with two known mixtures, the average deviation for five samples was ±0.3 percent. In the actual application, it was necessary to occasionally obtain the spectrum of a sample to show that no compound except the six analyzed for were present. This is good practice in all analyses where all variants are not positively known.

ACKNOWLEDGMENTS

The author acknowledges the assistance of G. M. Rassweiler, R. E. Nusbaum, and many other members of the Physics-Instrumentation and Organic Chemistry Departments of the Research Laboratories Division, General Motors Corporation.

Symposium on Fine Particles and Resolution

Chicago, Illinois, June 9, 10, 11, 1949

A symposium dealing with the properties and behavior of fine particles will be held June 9, 10, and 11 at the Stevens Hotel, Chicago, Illinois under the sponsorship of Armour Research Foundation of Illinois Institute of Technology and the Physics Department of the Institute.

Present plans indicate that the symposium will be composed of four sessions under the direction of scientists who are specialists on the individual topics. The sessions and session chairmen are:

"Light Scattering"—Dr. P. J. Debye, Cornell University
"Resolution"—Dr. C. W. Mason, Cornell University
"Formation and Size Distribution of Dispersoids"—Dr. H. F. Mark Polytechnic Institute of Brooklyn
"Microscopy of Fine Particles"—Dr. Robley C. Williams, University of Michigan

Each chairman will have a group of experts at his session who will present formal papers or semi-formal remarks. Time for appropriate discussion will be available with discussion sessions directed by the session chairman.

In addition to the regular sessions there will be instrument displays and a photographic exhibit excluding electron and light micrographs dealing with subjects that are related to the symposium topics.

The general chairmen for the symposium are W. C. McCrone and C. F. Tufts; further information may be had by contacting them at the Armour Research Foundation of Illinois Institute of Technology, 35 West 33rd Street, Chicago 16, Illinois.

Questions for general discussion at the sessions will be considered by the session chairmen if the questions are submitted in writing prior to the day the symposium convenes. Such questions should be addressed to the attention of the general chairmen.

Persons wishing to participate in the photographic exhibit should submit their material before June 6, 1949. Prints should be sent to the attention of the general chairmen.

The SYMPOSIUM ON FINE PARTICLES AND RESOLUTION is an outgrowth of the highly successful SYMPOSIUM ON ELECTRON AND LIGHT MICROSCOPY which was held in Chicago last year and attended by more than 250 scientists.

On the Prismatic Properties of Ophthalmic Lenses

GEORGE SAFONOV
Department of Physics, California Institute of Technology, Pasadena, California
AND
ROBERT GRAHAM
Director of Research, Plastic Optics Company, Pasadena, California

(Received August 30, 1948)

Equations are presented that locate a lens point which will direct light from an object of regard toward the eye's center. These equations connect the coordinates of the lens point with those of the object point and the various geometrical and optical parameters associated with a wearer of an arbitrary prescription involving spherical and cylindrical corrections. The three-dimensional consideration points out the existence of certain "surfaces of binocular vision." The text comments on the general impossibility of locating bifocal segments utilizing only the prismatic properties of the prescription and on a method to compute the magnitude of induced hyperphoria.

A graphical scheme to facilitate the use of the equations is included.

FOR each object point situated before an eye and its lens, there should be a lens point with that prismatic property necessary to guide a ray from the object toward the eye's center. The following employs computations compatible with what is generally termed "first-order theory." It is within the range of this theory to consider a thin lens as an equivalent continuum of prisms distributed over a plane.

Consider an equivalent lens plane normal to its optical axis. Let the eye be centered at point "O" of this axis (as shown in Fig. 1). Let the lens axis be the OZ axis of a Cartesian coordinate system with the origin at "O." The OX and OY axes are shown in the figure. The lens is centered at (O, O, e). An object point, "P," is located by (H, V, ϵ). The point of the lens "P'" which deflects a ray from P toward O has the coordinates (x, y, e).

Let F_s and F_c be the dioptral powers of sphere and cylinder associated with the lens. Let α^* be the angle between the cylinder axis and a line through (O, O, e) parallel to OX (see Fig. 1).

Applying the conditions that the magnitude and direction of prismatic deflection encountered at (x, y, e) due to F_s, F_c, and α are equal to the magnitude and direction computed from the geometrical quantities x, y, e, H, V, ϵ, there results two equations. Upon solving these two equations for x and y one obtains:

$$x=\gamma_x x_0, \tag{1}$$

$$y=\gamma_y y_0. \tag{2}$$

Here x_0 and y_0 are the values assumed by x, y if $F_c=0$. γ_x and γ_y are factors which modify x_0, y_0 to account for the existence of cylindrical correction. These quantities are defined as follows:

$$x_0=(e/\epsilon)\left(H/1-e\left(1-\frac{e}{\epsilon}\right)F_s\right), \tag{3}$$

$$y_0=(e/\epsilon)\left(V/1-e\left(1-\frac{e}{\epsilon}\right)F_s\right), \tag{4}$$

$$\gamma_x=\tfrac{1}{2}[1+\lambda+(1-\lambda)(\cos(2\alpha-K)/\cos K)], \tag{5}$$

$$\gamma_y=\tfrac{1}{2}[1+\lambda+(1-\lambda)(\sin(2\alpha-K)/\sin K)], \tag{6}$$

where

$$\lambda=\left(1-e\left(1-\frac{e}{\epsilon}\right)F_s\Big/1-e\left(1-\frac{e}{\epsilon}\right)(F_s+F_c)\right)$$

and

$$K=\tan^{-1}(V/H).$$

Since the center of a regular bifocal segment has zero prism, the placement of such a segment at x, y, e would not change the prismatic properties here. Indeed, considering a single eye, this would be the natural place to locate a segment for the purpose of scrutinizing points in the neighborhood of H, V, ϵ.

Consider now two distance lenses with their axes parallel. Let F_{s1}, F_{c1}, α_1 and F_{s2}, F_{c2}, α_2 denote the optical properties of the two lenses. The subscript 1 specifies the quantities associated with the lens worn before the right eye, while 2 accomplishes the same purpose for the left. Let the two lens axes be separated by a distance $2P$ which is the pupillary separation of the wearer. Thus the coordinates of the right eye and its lens center are (O, O, O) and (O, O, e), respectively. For the left eye, these coordinates are $(2P, O, O)$ and $(2P, O, e)$. With this convention established, Eqs. (1) to (6) may be used in their given form to compute x_1, y_1 in terms

* α is considered positive if the axis of cylinder lies below the plane $y=0$ on the nasal side of the lens ($-90° \leqslant \alpha \leqslant 90°$).

of e, H, V, ϵ and F_{s1}, F_{c1}, α_1. If it be understood that x_2 is the "horizontal inset" from $(2P, O, e)$ of the appropriate lens point P_2' used to view the object point (H, V, ϵ), the following define x_2, y_2:

$$x_2 = \gamma_{x2} x_{02}, \tag{7}$$

$$y_2 = \gamma_{y2} y_{02}, \tag{8}$$

where

$$x_{02} = \frac{e}{\epsilon} \frac{(2P-H)}{1-e[(1-(e/\epsilon)]F_{s2}}, \tag{9}$$

$$y_{02} = \frac{e}{\epsilon} \frac{V}{1-e[1-(e/\epsilon)]F_{s2}}, \tag{10}$$

$$\gamma_{x2} = \frac{1}{2}\left[1+\lambda_2+(1-\lambda_2)\frac{\cos(2\alpha_2-K_2)}{\cos K_2}\right], \tag{11}$$

$$\gamma_{y2} = \frac{1}{2}\left[1+\lambda_2+(1-\lambda_2)\frac{\sin(2\alpha_2-K_2)}{\sin K_2}\right], \tag{12}$$

where

$$\lambda_2 = \frac{1-e[1-(e/\epsilon)]F_{s2}}{1-e[1-(e/\epsilon)](F_{s2}+F_{c2})}$$

and

$$K_2 = \tan^{-1}(V/2P-H).$$

It is clear that y_1 will not generally equal y_2 for a randomly chosen object point H, V, ϵ. However, for normal vision y_1 must be equal to y_2, for otherwise the visual axes of the two eyes would not remain coplanar. If there exist any points H, V, ϵ which may be viewed subject to the restriction $y_1 = y_2$, they must lie on the surface defined by

$$\gamma_{y1} y_{01} = \gamma_{y2} y_{02}. \tag{13}$$

The surface** whose analytical equation is (13) might be appropriately termed the "surface of binocular vision" associated with the wearer's prescription. It is to be noted that this surface is defined with respect to the head as a reference frame. Thus the wearer is not left binocularly blind of any points he wishes to view; it is simply a matter of turning the head so as to make his particular surface include the desired object point.

An especially simple case is the prescription $F_{s1} \neq F_{s2}$, $F_{c1} = F_{c2} = 0$. Here the binocular surface is $V = 0$—that is, the plane determined by the two lens axes.

The binocular surface does not always include the points of the median plane, $H = P$, normally occupied by reading matter. Thus it is not always possible to locate bifocal segments utilizing only the prism intrinsic to the distance lenses.

The magnitude of induced hyperphoria encountered while viewing a general object point (H, V, ϵ) may be computed from Eqs. (2) and (8). A measure of this quantity is given by $(y_2 - y_1/e)$.

It is interesting that the somewhat laborious calculation of x, y indicated by Eqs. (1) through (6) may be replaced by use of a simple graphical scheme. For a given set of values of e, ϵ, (V/H), one may readily determine x, y in terms of an arbitrary prescription specified by F_s, F_c, α, and H. (H would generally be the semipupillary distance in the case of a bifocal inset computation.) Figure 2 is a curve sheet constructed to determine x when the following values are assumed:

$e = 25$ mm,
$\epsilon = 42.5$ cm (corresponding to a "reading distance" of $\epsilon - e = 40$ cm),

$V/H = 4.52$.

Figure 2 shows four families of curves drawn in four quadrants. To exemplify the use of this curve sheet, consider the prescription $F_s = 10$ diopters, $F_c = -5$ diopters, $\alpha = 30°$, and $H = 30$ mm. Beginning a rectangular construction at point ① which corresponds to the power of sphere, a line is drawn upward to point ② as shown. ② marks the intersection of this line with a curve labeled $H = 30$ mm. A line is drawn from ② toward the right to point ③. Point ③ is the intersection of this second construction line with the scale denoting "horizontal inset for zero cylinder." Thus if the prescription were void of cylindrical correction we would find $x_0 = 2.31$ mm; this would complete the determination of x.

In the example cited, however, $F_c = -5$ diopters. Thus the line ⟨②①⟩ is extended downwardly to point ④. ④ corresponds to the intersection of this line with a member of the family of curves shown in the third quadrant that would be labeled $F_c = -5$.

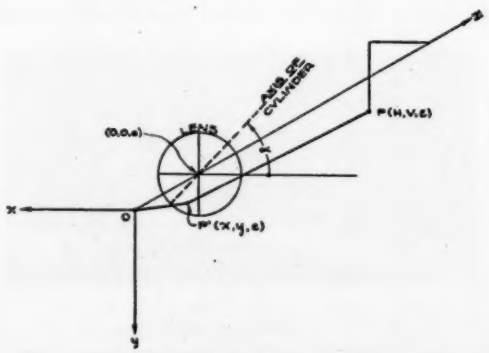

FIG. 1.

** In certain cases all points are available to normal binocular vision (e.g., $F_{s1} = F_{s2}$ and $F_{c1} = F_{c2} = 0$) and (13) does not define a surface.

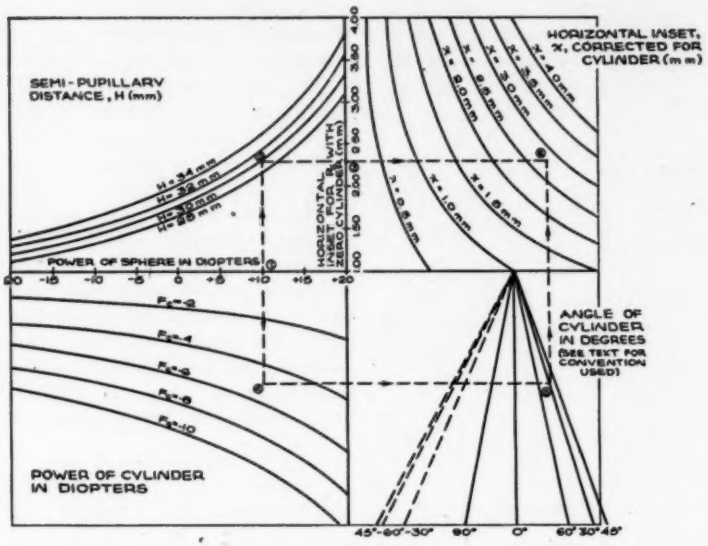

FIG. 2.

diopters. From ④ a line is drawn to the right to point ⑤ where it intersects the curve marked $\alpha = 30°$. Lastly, from ⑤ a line is drawn upward to point ⑥. ⑥ is the intersection of this line with line ⟨②③⟩ extended to the right. ⑥ is found to lie in the neighborhood of a curve that would have been labeled $x = 2.8$ mm; this is the value desired. The graphical determination is much more rapid than a straight computation involving Eqs. (1) through (6) which may be used to check the above result.

A similar curve sheet may be constructed for the determination of y.

The authors wish to express their appreciation to Professor Glenn Fry of Ohio State University for his work in checking the equations of the text and for his many helpful suggestions.

Letters to the Editor

The Green Segment Seen from an Airplane

E. O. HULBURT
Naval Research Laboratory, Washington, D. C.
March 10, 1949

THE green color of the disappearing rim of the setting sun which may occasionally be seen has been well described.[1] It is due to the prismatic dispersion of the rays of sunlight by the normal atmosphere, and sometimes is accentuated in conditions of mirage when abnormal atmospheric refraction occurs. An unusually clear atmosphere is necessary, such that the sun is yellowish, and not red, as it begins to sink below the horizon. A red setting sun means that the blue and green portions of the spectrum are relatively strongly attenuated by the atmosphere, and hence indicates that conditions are not favorable for seeing a greenish segment.

From an airplane flying at sunset over Tennessee on February 6, 1949, the green segment was clearly seen. The plane was flying at 8000 feet over a complete cloud cover about 2000 feet below. The cloud horizon was smooth and sharp, and the air was very clear down to the horizon. The possibility of seeing the green segment had occurred to me, and I began observations by using one eye to watch the sun and keeping the other eye screened so that it might be ready for use without possible trouble from after images. This turned out to be unnecessary because by continually shifting the gaze no disturbing after images were seen. The sun appeared a pure yellow as it began to dip below the cloud-sky horizon. Just before the slim remaining segment disappeared it turned from yellow to white and finally to a distinct green; the color was pale green and not a vivid emerald green. The entire phenomenon lasted less than three and more than two seconds. Its duration seemed very brief and I am not surprised that earlier observers have spoken of it as a green "flash." If what I saw was a fair example of the green segment I was left with the impression that the phenomenon was rather fleeting and not very impressive, and that it would usually escape casual observation.

The prismatic dispersion of the normal atmosphere for a ray tangent to the surface of the earth is 22 seconds of arc between wave-lengths H_α and H_β, as calculated after the manner of Rayleigh;[2] and at 6000 feet altitude, where the pressure is about 0.8 atmosphere. It amounts to about 18 seconds of arc. The observed time of about 2 seconds which elapsed from the appearance of the green segment to its disappearance below the horizon corresponded to an angular motion of the sun of about 23 seconds of arc. This was near enough to the calculated value 18 seconds of arc to indicate that the dispersion of the normal atmosphere was adequate to account for the observed green segment.

[1] M. Minnaert, *Light and Color in the Open Air* (G. Bell and Sons, London, 1940), pages 58–62.
[2] Rayleigh, Proc. Roy. Soc. **126**, 311–318 (1930).

The Reflecting Power of Glass with a Thin and with a Thick Film

ANTONIN VAŠIČEK
The Beneš Technical University, Brno, Czechoslovakia
February 25, 1949

CONSIDERING multiple reflections of light on the film the formulas for the reflecting power of glass with a *thin* and with a *thick* film are deduced in general separately for both kinds of film. The reflecting power of glass having a *thick* film is given by the formula

$$\rho=1-\frac{4n_1n}{(n+1)(n_1^2+n)}=1-J, \tag{1}$$

and similarly for the reflecting power of glass with a *thin* film we get

$$\rho'=1-\frac{8n_1^2n}{2n_1^2(n+1)^2-(n_1^2-1)(n^2-n_1^2)+(n_1^2-1)(n^2-n_1^2)\cos x}, \tag{2}$$

where n_1 and n are the refractive indices of the film and of the glass, x is the path difference in the thin film.

Assuming that the reflection of light from glass with a *thick* film is a special case as mean value of the reflection of light from glass with a *thin* film we can deduce the formula (1) from Eq. (2) in the following way.

Substituting in Eq. (2),

$$\begin{aligned}2n_1^2(n+1)^2&=p,\\(n_1^2-1)(n^2-n_1^2)&=q,\end{aligned} \tag{3}$$

we get

$$\rho'=1-\frac{8n_1^2n}{p-q+q\cos x}. \tag{2'}$$

Then we calculate the mean value of the integral

$$J=\frac{1}{2\pi}\int_0^{2\pi}\frac{8n_1^2n}{p-q+q\cos x}dx=\frac{8n_1^2n}{p-q}\frac{1}{2\pi}\int_0^{2\pi}\frac{dx}{1+\epsilon\cos x}, \tag{4}$$

where

$$\epsilon=q/(p-q). \tag{5}$$

Having respect to the solving of integrals,

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{dx}{1+\epsilon\cos x}=\frac{1}{(1-\epsilon^2)^{\frac{1}{2}}}=\frac{p-q}{(p^2-2pq)^{\frac{1}{2}}}, \tag{6}$$

we have then

$$J=\frac{8n_1^2n}{(p^2-2pq)^{\frac{1}{2}}}=\frac{4n_1n}{(n+1)(n_1^2+n)}, \tag{7}$$

in accordance with Eq. (1) for the reflection of light from glass with a *thick* film.

Combustion Studies with the Orthicon Spectrograph

J. T. AGNEW, R. G. FRANKLIN, R. E. BENN, AND A. BAZARIAN
Franklin Institute Laboratories for Research and Development, Philadelphia, Pennsylvania
March 10, 1949

APPLICATION of the image orthicon and electronic instrumentation to spectroscopy* has provided increased light sensitivity, greater spectral range of response, and time resolution of a higher order than can be obtained with a moving film or plate. The instrument is ideal for studying transient

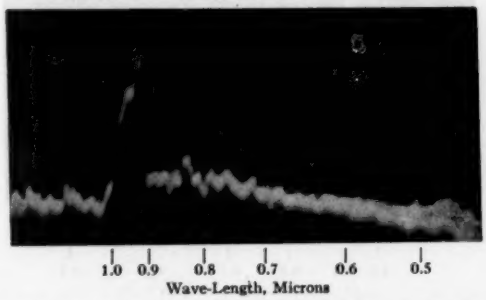

FIG. 1. Radiant energy emission during one millisecond interval from 10th to 11th millisecond.

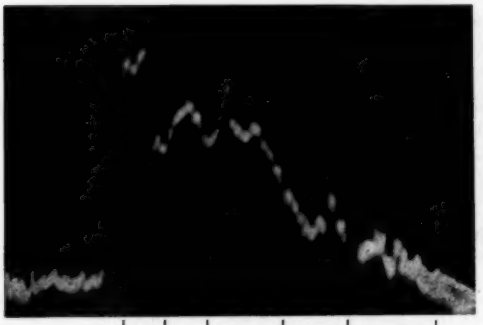

FIG. 2. Radiant energy emission during one millisecond interval from 14th to 15th millisecond.

or rapidly varying light sources and has already proved useful in a preliminary combustion study.

The instrument consists essentially of a television pick-up tube, the RCA image orthicon, whose photo-sensitive surface is placed in the focal plane of a large-aperture spectrograph. A curve of intensity *vs.* wave-length, at any chosen instant or succession of instants, appears on the screen of an oscilloscope whose scans are synchronized with those of the orthicon. Recurrent traces are photographed with a high speed drum camera.

A recent study has been made of the explosion in a bomb of a mixture consisting of 3.5 percent *n*-heptane, 62.8 percent oxygen, and 33.7 percent nitrogen, by volume. This can be used as an illustration of the advantages of the orthicon assembly. Single sweep scans were made at the rate of one every millisecond. They were photographed on a rotating-drum camera whose speed was such that successive traces were conveniently separated. Each scan represents the integrated intensity of the preceding millisecond.

A few representative traces from the series have been reproduced in the accompanying figures. From these and the original traces, a few general observations are apparent.

The growth with time of various emission lines and bands is clearly visible. Characteristic radiation first appeared in the vicinity of 0.93 to 0.96 microns (see Fig. 1). Red and yellow bands have appeared in 15 milliseconds (Fig. 2) and strong green bands in 25 milliseconds (Fig. 3).

The appearance of the bands and lines in the order red to green and their decay in the order green to red strongly suggests the important role of temperature in these phenomena. For the mixture studied, it appeared that the temperature increased rapidly during the first part of the combustion, reached a maximum which was maintained without great change for a time, and then gradually decreased. Paralleling the temperature changes was the appearance of the red bands and lines and then the addition of the other spectral regions, the green appearing only at maximum temperature and decaying after the temperature had again fallen below that necessary to excite the region.

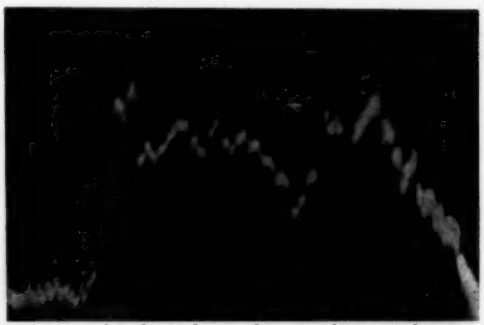

FIG. 3. Radiant energy emission during one millisecond interval from 24th to 25th millisecond.

It might be pointed out that with a single explosion, data comprising the whole spectral region from 0.45 to 1.0 micron is obtained, and some bands are visible which might possibly escape detection if a standard photographic method were used. Also, the feature of obtaining complete spectra of the explosion at a rate of one every millisecond is unique due to the high sensitivity and storage feature of the orthicon tube.

It is also important to note that the rate of one trace per millisecond was arbitrarily chosen for this particular explosion. Faster rates could be used if the source were of shorter duration, particularly so if it were more intense. The maximum rate, using commercial components, would be approximately 15,750 traces per second, in which case the storage time would be equal to the tracing time of approximately 63 microseconds. The actual time consumed in producing the traces shown in Figs. 1–3 was 63 microseconds while the storage time was 1 millisecond.

* Benn, Foote, and Chase, *The Image Orthicon in Spectroscopy*, to be published.

On the Relative Measurement of Kerr Constants

HAROLD A. SCHERAGA,* MARCUS E. HOBBS, AND PAUL M. GROSS
Department of Chemistry, Duke University, Durham, North Carolina
March 7, 1949

RELATIVE measurements of Kerr constants can be made by using two Kerr cells, wherein the phase difference introduced in the light beam by the first cell is compensated by an equal and opposite phase difference produced by the second cell.[1] From Kerr's law the following relation holds at the point of compensation:

$$\frac{B_1}{B_2}=\frac{l_2}{l_1}\cdot\frac{a_1^2}{a_2^2}\cdot\frac{V_2^2}{V_1^2}$$

where the subscript 1 refers to the first, or standard cell and the subscript 2 refers to the second, or measuring cell. B is the Kerr constant of the liquid, l and a are the plate lengths and separation, respectively, and V is the voltage applied to the plates.

In using this method high potentials were applied to the plates of each cell in such a manner that the voltages on each cell were in phase at all times. This was accomplished by feeding the primary of each of two equivalent, but not carefully matched, 110–10,000 volt Thordarson transformers from the same power source. Variations in the voltages impressed on the primaries of the transformers were effected by means of equivalent non-inductive slide wire resistors. The voltages so applied were read on the same voltmeter by use of a double pole double throw switch connecting the voltmeter with the respective primaries of the two transformers. The secondary of each transformer was shunted with a fixed "bleeder" resistance the value of which was such that most of the current in the secondary was carried by this resistance thus stabilizing the secondary circuit. These resistances served the further purpose of minimizing effects caused by the reactance of the Kerr cells.

With the above arrangement of the apparatus it may be safely assumed that

$$V_{1s(t)}=K_1V_{1p(t)}$$
$$V_{2s(t)}=K_2V_{2p(t)}$$

where $K_1\simeq K_2$. The subscripts have the following significance: 1 and 2 refer to the system for the standard and measuring cells respectively, s and p refer to the secondary and primaries of the appropriate transformers and t is time at which the voltage is impressed and will be equivalent in both cases since phase shifting components have been largely eliminated or balanced. The small difference in time of travel of light between the two cells is negligible for the frequency used (60 c.p.s.).

Any differences in K_1 and K_2 and the separations and lengths of the cells can be absorbed in an apparatus constant, A, as given by the relation

$$B_1/B_2=A(V_{2p}^2/V_{1p}^2)$$

The apparatus constant was determined by filling both cells with the same liquid (so that $B_1=B_2$) and determining the voltage ratio over a range of potentials. This constant, A, was then used for subsequent measurements on other liquids.

Using sodium D light, two calibration runs with purified nitrobenzene and two with purified carbon disulfide gave a value of A equal to 0.90 ± 0.02. It will be noted that since both cells are filled with the same sample of liquid and since the characteristics of the two systems are essentially the same, A will be insensitive to traces of impurity in the liquid.

The Kerr constant of purified chlorobenzene relative to both nitrobenzene and carbon disulfide was determined, making use of the apparatus constant obtained from the calibration. The Kerr cells were not thermostatted; however, the temperature was determined at the time the voltage was measured so that the temperature dependence of the Kerr constant could be taken into account in making a comparison of the results obtained with those reported in the literature. Where the literature values were given for wave-lengths other than the sodium D line, dispersion corrections were made using Havelock's law.

The ratio of the Kerr constants of nitrobenzene and chlorobenzene was found to be 18.6 ± 1.0 as compared to a literature value of 18.0 using Grodde's data for chlorobenzene[2] and Ilberg's data[3] together with Gabler and Sokob's temperature coefficient[4] for nitrobenzene. The large variation in the reported Kerr constants for nitrobenzene should be noted.[3,4]

The ratio of the Kerr constants of carbon disulfide and chlorobenzene was found to be 0.259 ± 0.009 as compared with 0.240 using Bergholm's data[6] for carbon disulfide.

From the precision obtained in these preliminary experimental results on both the calibrations and the determination of the ratio of the Kerr constants, the maximum experimental error with this apparatus, constructed to test the validity of the method reported here, is estimated to be of the order of 5 percent. It is felt that, with a more refined apparatus designed to eliminate the usual errors in electrical birefringence measurements, this method is capable of much better precision and affords a considerable simplification in equipment design over that normally employed.

* Present address: Department of Chemistry, Cornell University Ithaca, New York.

[1] A review of the Kerr effect is given by H. A. Stuart, Hand. u. Jahrb. d. Chem. Phys. 10, 27 (1939).
[2] K. H. Grodde, Physik. Zeits. 39, 772 (1938).
[3] W. Ilberg, 29, 670 (1928).
[4] F. Gabler and P. Sokob, Zeits. f. Techn. Physik 17, 197 (1936).
[6] C. Bergholm, Ann. d. Physik 65, 128 (1921).

Book Review

Atomic Energy Levels—As Derived from the Analysis of Optical Spectra

VOL. I, SECTION 1, Circular No. 467.

This 75 page pamphlet is the first of a series being prepared at the National Bureau of Standards as part of a program on the critical compilation of all energy levels derived from observations of atomic spectra. The collected results will finally appear in bound volumes of approximately 500 pages each, made up of these sections.

This first section, written by Charlotte E. Moore, contains 55 tables of energy levels compiled from all of the known work on the atomic spectra generally designated as H*I*, D*I*, He*I*, He*II*, Li*I*, Li*II*, Li*III*, Be*I*–Be*IV*, B*I*–B*V*, C*I*–C*VI*, N*I*–N*VII*, O*I*–O*VIII*, and F*I*–F*VIII*. These results extend the analyses of these elements compiled and published 15 years ago by R. F. Bacher and S. Goudsmit under the title "Atomic Energy States."

Energy level values are given for the *normal states* and several of the lowest *excited states* of the highly ionized atoms of C, N, O, and F. The isoelectronic sequences starting with carbon, nitrogen, and oxygen have now been carried to their limits. In these limiting spectra the respective atoms have been stripped of all but the one remaining electron needed to carry out transitions. In the most highly ionized state of fluorine listed all but two electrons have been removed.

This extremely useful set of tables will be gladly welcomed by many scientific groups, not only those doing research in atomic physics but those in the field of nuclear research as well. To those involved in the compilation of these tables we can say congratulations and thanks for a job well begun.

HARVEY E. WHITE
University of California

The Birefringent Filter: A Correction

[J. Opt. Soc. Am. 39, 229 (1949)]

JOHN W. EVANS
Harvard Observatory, Harvard University, Cambridge Massachusetts

SHORTLY after this paper went to press, the author was greatly embarrassed to discover an error in his remarks concerning the field of the split element filter. The field, is, in general, approximately the same as that of the equivalent simple filter. Within this limitation, however, the split element filter performs satisfactorily, and for some purposes the reduction in the number of polarizers required is important.

If, using the notation of the original paper, we let

$$\begin{aligned} n_m - n_q &= \delta \\ n_m + n_q &= n_j, \end{aligned} \tag{1}$$

Eqs. (V.8) reduce to

$$\begin{aligned} \tau_r &= 1 - \cos^2\pi n_p \sin^2\pi\delta - \sin^2\pi n_j \sin^2\pi n_p \\ \tau_s &= \cos^2\pi n_p \sin^2\pi\delta + \sin^2\pi n_j \sin^2\pi n_p. \end{aligned} \tag{2}$$

The filter is constructed with n_m and n_q equal for light parallel to the instrumental axis ($\phi=0$). For light inclined to this axis, however, they are no longer equal, and, from Eq. (IV.2),

$$\delta = \frac{n_j}{2} k\phi^2 \left(\frac{1}{\gamma} + \frac{1}{\alpha}\right) \cos 2\theta. \tag{3}$$

There is, therefore, a system of hyperbolic fringes (composed of light in the wave-length intervals between the transmission maxima of the p component) superposed on a field of broad circular fringes indicated by Eq. (IV.1). If n_p is approximately $\frac{1}{2}$ over the spectral region considered, we have essentially Lyot's first type filter, and the fringes are very weak. The field of the split element filter can be increased by constructing the two components of the split elements individually in any one of Lyot's three wide-field forms. Since the fringes are then circular and identical in each of the two components, $\delta=0$ for all values of ϕ. The off axis effects are then given by the appropriate equations of section IV of the original paper.

Since no gain in field results from crossing the γ-axes of the m and q components, it is pertinent to ask whether this construction is the most advantageous. The transmission of a split element unit with the γ-axes of the m and q components parallel readily follows from Eq. (V.6) of the original paper if we interchange α and γ as multipliers of d_q/λ. We find

$$\begin{aligned} \tau_r &= \cos^2\pi(n_m - n_q) - \sin 2\pi n_n \sin 2\pi n_q \cos^2\pi n_p \\ \tau_s &= \sin^2\pi(n_m - n_q) + \sin 2\pi n_m \sin 2\pi n_q \cos^2\pi n_p. \end{aligned} \tag{4}$$

Expressed in terms of δ and n_j, this becomes:

$$\begin{aligned} \tau_r &= 1 - \sin^2\pi n_p \sin^2\pi\delta - \sin^2\pi n_j \cos^2\pi n_p \\ \tau_s &= \sin^2\pi n_p \sin^2\pi\delta + \sin^2\pi n_j \cos^2\pi n_p. \end{aligned} \tag{5}$$

If δ is zero for light parallel to the instrumental axis, it remains zero at all values of ϕ. The field is the same as that of the equivalent simple filter. The expression for τ_s is then

$$\tau_s = \sin^2\pi n_j \cos^2\pi n_p.$$

This arrangement does offer some advantage over the split element unit with the γ-axes of the m and q components crossed. A filter of l birefringent elements can be constructed with its first $l/2$ elements (i.e., the thin elements) sandwiched between the halves of the $l/2$ thick elements. Since the transmission of the middle element in each unit is $\cos^2\pi n$, the n-values are integral at the desired wave-length, and can be made proportional to the powers of 2. The $l/2$ thin elements therefore function equally well throughout the spectrum. Since the burden of suppressing the light at wave-lengths far removed from those of the transmission bands falls mainly on these thinner elements, this property is a real advantage.